NIBR WORKING PAPER 2022:101

NORWEGIAN INSTITUTE FOR URBAN AND REGIONAL RESEARCH NIBR

What is digital social science?

Challenges and opportunities of digitalization for the social sciences

Henrik Wiig and Yuri Kasahara

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Henrik Wiig
Yuri Kasahara

What is digital social science?

Challenges and opportunities of digitalization for the social sciences

NIBR WORKING PAPER 2022:101

Title:	What is digital social science? Challenges and opportunities of digitalization for the social sciences
Author:	Henrik Wiig and Yuri Kasahara
NIBR Working paper:	2022:101
ISSN:	0801-1702
ISBN:	978-82-8309-362-9
Project name:	Digital social science strategy
Financial supporter:	Internal funding
Head of project:	Henrik Wiig
Abstract:	Information collection and registering of data is today digital and stored on computers. Advanced statistical models analyse this information to recommend actions. The whole cognitive process can take place in computers rather than by humans. How does such artificial intelligence (AI) systems and machine learning (ML) influence society and how can social researchers apply these new methodological tools in their ongoing research? The new academic field of digital social science search responds to both challenges. This report describes how ML and AI functions, benefits and possible adverse challenges, both in society in general and in social science research in specific. We identify the most common uses of ML and AI in social science, and furthermore summarizes some literature on their effect on society and our human lives. Finally, we also make some recommendations on how the research institutes in the Centre for welfare and labour research at Oslo Metropolitan University can implement a program of digital social science.
Date:	February 2022
Pages:	39
Publisher:	Norwegian Institute for Urban and Regional Research OsloMet – Oslo Metropolitan University Postboks 4 St. Olavs plass 0130 OSLO Telephone: (+47) 67 23 50 00 E-mail: post-nibr@oslomet.no http://www.oslomet.no/nibr

Preface

This working paper has been written by a team consisting of Henrik Wiig and Yuri Kasahara. Both are senior researcher at the Department for International Studies and Migration at the Norwegian Institute for Urban and Regional Research (NIBR) at Oslo Metropolitan University.

The working paper is commissioned and funded by NIBR as a first approach to formulate a strategy for the institute to develop our own approach to digital social science.

Oslo, March 2022

Kristian Tronstad
Head of Research

Table of Contents

Preface 1
Summary 3
1 Introduction and motivation 5
2 The flood of (unstructured) data 7
2.1 Big data 7
2.2 Structuring text data: an example 8
3 Machine learning role in analysing structured datasets 10
3.1 Predict rather than explain 10
3.2 Function and types of ML 10
3.3 From advanced regressions to neural networks and deep learning 13
4 Automated decisions finalize AI systems 18
5 Application in social sciences 19
5.1 Accessing and structuring data 19
5.2 Coding and Classification – Easier and more insightful analysis 20
5.3 Social Networks 22
5.4 Physical geography and spatial analysis 22
5.5 Construct explanatory variables for causal analysis 23
6 Application in public administration 25
7 Research on the effect of digitalization on society 27
8 Challenges in digital social science 30
8.1 Expertise needed 30
8.2 Ethics and biases 31
8.2.1 Introduction 31
8.2.2 Access to data and individual privacy 32
8.2.3 Justice and fairness 33
9 Conclusion 36
References 37

Summary

Human beings observe, analyse and then take actions. Today, all steps in human cognition processes can be undertaken by machines in *artificial intelligence (AI) systems*. Human actions and results, historically and in real time, can be recorded digitally through keystrokes on a computer, surveillance cameras, satellites, and records of trade for institutional and well as individual actions. However, such *big data* are organized in many ways, and require sorting and structuring to be analysed. All such digital information, up to the time of taking action, is 'observed' by a machine: this constitutes the description of reality on which the machine can take action. New statistical methods of *machine learning* (ML) identify states of the world as well as optimal actions to be taken given these circumstances, based on historical experience. Computers can implement such predetermined systems as 'take action A, if state 1; action B, if state 2…', etc. Such AI systems can replace humans throughout the decision-making chain, but humans still play important roles in two central functions: it is humans who categorize observations to be analysed by ML, and who evaluate the usefulness of the outcomes of the various actions available at each stage.

The academic field of *digital social science* applies these new analytical techniques and sources of data, as well as studies how digitalization with AI systems and ML affects how we live and the societies we form. The present report describes various ML approaches and how they can be used in social research. Researchers can classify observations though *supervised ML,* and thereby analyse quantitatively large amounts of different types of data. Through *unsupervised ML* for grouping elements we can identify known categories and discover new ones in observations that inform us about society. Using *reinforcement ML* we can optimize responses in repeated interaction between humans or nature, to identify optimal strategies in institution building, business development, etc. The first generation of *algorithms* in ML uses regression techniques in various form. The increase in data storage capacity and the speed of the data processors over the past decade has now enabled numerical simulation of *neural networks* and *deep learning* that produce more precise predictions and hence also more accurate automated actions.

These new sources of data and analytical tools are useful in many types of research. Text analysis, where identification and categorization to study dynamics over time and across actors can now be done on whole libraries: this had previously been limited to the amount of text that researchers and their assistants could classify by reading themselves. Today it is sufficient to read and classify some examples, after which the machine can repeat the structure, applying it to any text made available. Of special interest here is network analysis through tracing interactions on the Internet and geographic positioning through GPS to enable area planning.

We can observe the rapid introduction of digitalized information and AI systems in all parts of society. Public administrations render digital services to the citizenry, and introduce automated administrative procedures. Early warning systems that enable early interventions, e.g. predicting failure on exams or payment of financial obligations based on digital information on past behaviour, can be used to prevent that a bad outcome becomes a fact.

Society and the economy adapt to the possibilities offered by digitalization and AI systems. Economists study how markets change and create new business models – leading to market concentration for some large companies, but also helping small niche actors to be noticed by consumers through the Internet. However, the increased individualization of the labour market through the 'gig-economy' is a notable change in

the overall functioning of the market economy. Sociologists are studying generational behaviour changes, for instance, how teenagers have more close friends today and wider networks in spite of spending more time at home.

Like any quantitative analytical method, ML is based on statistical description of groups. The resultant actions are hence based on the perceived stereotypes rather than the (unknown) specific features of the individual affected by that action. In turn, this can lead to discrimination and irrelevant categorization. Therefore, this report concludes that it is vital that social scientists with an in-depth understanding of society should be involved in ML and AI projects, to create trustworthy and transparent systems.

1 Introduction and motivation

Today it is a truism to say that we live in a digital world. We human beings leave digital traces from most of our actions. Interactions in the social media, buying groceries at the local store, work, taking a picture, going for a walk with your phone in your pocket, or being filmed by surveillance camera: anything that involves a keystroke or an interaction with a digital device is recorded, generating data that can be stored and later analysed.

The expansion of digital devices and spaces created by them (digital realities) have revolutionized how people interact. We now meet virtually, whereas effective communication had previously required meeting physically. As the recent pandemic has shown, activities such as teaching, or even medical examinations, can be conducted remotely through digital media. Digitization offers not only new sources of data,[1] but also new ways of interacting, whose consequences for individuals and society we still do not fully grasp.

As this digital revolution takes hold of our lives, we begin to realize that these new and abundant sources of data can be analysed to understand and predict individual and social behaviour, as well as to automatize activities previously conducted by humans.

To a considerable extent, this process has been possible due to increased data storage capacity, faster data processors and parallel developments in computer science, whereby new quantitative methodologies have made possible the processing and analysis of enormous amounts of digital information in the emerging field of *data science.*[2] Companies use these new data and analytical methods to increase their productivity and profits. Governments improve services, tax collection, etc. through evidence-based policies, and – also in negative and dangerous ways – to strengthen political and social control.

Social scientists are catching up and are now analysing these new data sources and methodologies in their work. The fast-growing field of *digital* (or *computational*) *social science*[3] explores how we can structure these data and analyse them by applying new quantitative methods, conduct our research more efficiently, and opens for quantitative analysis of sectors and fields that previously lacked relevant data.[4]

Another important aspect involves focusing on the societal effects of digitization. Thus, digital social science is commonly defined as research on the social aspects and

---

[1] Previously, a medical surgery operation would record only the data required by regulations, e.g. patient, doctor, diagnosis and outcome. Today, if the full operation is filmed, instructions and movements will be recorded, yielding data on many details and thereby making possible quantitative analysis to identify what leads to successful operations – data that later can be included as recommendations or requirements for future operations.

[2] It has been estimated that the world produced 40 zetabytes of data only in 2020, or 1.7MB per second for every person in the world: https://www.ibm.com/blogs/journey-to-ai/2020/06/netezza-and-ibm-cloud-pak-a-knockout-combo-for-tough-data/

[3] The use of these two terms refers more to the disciplinary background of the researcher than any substantive difference between them. The older term, *computational social sciences,* originated in academic communities interested in developing simulation models for human behaviour (e.g. agent-based models) or analysing big datasets (see Edelman et al., 2020). The more recent term *digital social sciences* has been adopted by academics with roots in the humanities and disciplines like anthropology and sociology (Jemielniak, 2020). Further, some institutions (among them, the London School of Economics and the University of Oxford) have started using the term *social data science* to define the field.

[4] Briefly: although all information can be represented in a digital way, i.e. structures of bits of 0 and 1 in a computer, we restrict our definition of *digital social science* to (i) process of collecting and organizing unstructured data and/or (ii) applying newly developed quantitative methods (ML) for structured data. Thus, we do not focus on traditional causal inference methods on structured data in the concept.

impacts of digital technologies, as well as the actual application of such technologies in research methodologies (Ignatow and Robinson, 2017).

However, today a third element needs to be considered. The Norwegian public administration has shown increasing interest in digital social sciences, by developing new administrative tools that use machine learning (ML). Many traditional funders of applied social science research, such as municipal administrations and national agencies (e.g. BufDir, IMDi, NAV) are interested in commissioning social science institutes to leverage this potential to improve their analyses as a basis for policies and the implementation and evaluation of policy designs. The digital social sciences provide perspectives that individual data scientists alone cannot offer in terms of thematic and empirical knowledge.

> The public sector and AI: NAV has its own Artificial Intelligence Lab for developing tools and analyses based on machine learning. BufDir, with participation of NIBR, has funded a project to develop an automated classification model for hate speech against Muslims. IMDi is currently funding a project aimed at developing an optimization algorithm for identifying municipalities where asylum-seekers have the highest probability of getting employment.

Taking these three elements as a basis, this report systematically presents various facets of the field of digital social sciences, with emphasis on the application of digital technologies to research methodologies. This, it is hoped, will enable researchers to tap into the potential of new data sources, methods, and topics of research by incorporating a digital social sciences perspective in their work.

In a more pragmatic perspective, searching for synergies with the digital social sciences marks an important step towards offering new analyses to clients and innovative research projects to funding agencies. Regardless, social scientists cannot afford to ignore digital social sciences if they want to remain relevant on the research frontier in the future.

This report starts with a description of how the massive inflow of empirical information/ data challenges social science. We then turn to the role of ML in structuring the data inflow. This is followed by a brief account of how Artificial Intelligence (AI) encapsulates the entire process of collecting, analysing, and acting upon data. After this introduction to digital technologies and programmes we discuss the application of the methods of ML and AI within the social sciences, as well as the application within public administration. We conclude by noting some obstacles to building applied digital social science capacity, and the ethical challenges involved.

2 The flood of (unstructured) data

A narrow definition of ‘digital social science’ sees it as a set of methods and techniques for collecting, structuring and analysing data for the purposes of social science research.

2.1 Big data

In the beginning there was data. Regardless of how we define data or how we conduct data collection, all social scientists use data in their analyses. Also applied theoretical work requires data, for testing its empirical validity. Social scientists trained in qualitative methods use in-depth interviews, participatory observations and texts as data sources. Researchers trained in quantitative methods use surveys and official statistics as classical sources of data.

Another way to categorize types of data concerns the level of organization or structure. By ‘structure’ is meant how accessible specific information is: thus, a dataset organized by variables (columns) and units of analysis (rows) is a *highly structured* dataset. A researcher interested in specific information contained in the dataset can easily access and use the data in a meaningful way, if there is adequate documentation of what each variable in the dataset represents. By contrast, a collection of texts (e.g. comments from the social media) can be considered as an example of *unstructured* data. Texts and other unstructured data require processing (categorization/classification) in order to be analysed with quantitative methods.

Our times are characterized by the availability of massive amounts of unstructured data. The term *big data* is often applied to such collections of unstructured data. Digital images – from social media to medical exams, audio files, sensor data (including GPS data) – are common examples of big (i.e. unstructured) data. However, in our use of the term, the adjective ‘big’ in big data refers not to size, but more to the need to structure the data in line with our definition of the concept of 'digital social science’ as such.

That being said, however, there is no clear definition of ‘big data’ in the literature. Favaretto et al. (2020) interviewed 39 big-data experts to identify common features in their definitions of the term. They found that the Vs – *volume* (huge amounts), *velocity* (high-speed processing) and *variety* (heterogeneous data) – were the traditional attributes of big data. Three new Vs have now been added to the list: *veracity* (truthfulness), *value* (usefulness) and *variability* (high dimensionality). Favaretto and colleagues (2020) further note the US National Science Foundation definition of big data as ‘large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, internet transactions, email, video, click streams, and/or all other digital sources available today and in the future’. They also find that the concept often includes the quantitative techniques applied to structure and analyse such data.

Based on this definition of big data, we can note that the *first aspect* of what we define as ‘digital social sciences’ concerns the set of methods and techniques for collecting, structuring and analysing big data for the purposes of social science research. From a digital social sciences perspective, the traditional categorization into qualitative and quantitative data sources becomes blurred, as all data can be digitized. Interviews can be transcribed in digital format. Fieldwork notes or pictures are now digital. Wherever the data collected can be digitized, applying digital social sciences methods is possible:

thus, also qualitative information can be transferred and interpreted to become quantitative information, open for analysis through quantitative methods. In comparison to traditional sources of data, these new sources are more complex and too voluminous to be structured and analysed by traditional methods or procedures. Indeed, the digital social sciences also aim to incorporate the methods developed by computer scientists in the fields of *machine learning* (ML) and *artificial intelligence* (AI) in the social sciences, to assist in making sense of these new unstructured data sources. The *second aspect* is simply to study how digitalization, ML and AI affect society through applying traditional or the new social research methods.

Viewed schematically, the processing of new digital information takes place in three stages. First, the data must be collected and structured into a matrix suitable for quantitative analysis. Beyond the field of the social, this stage is often the realm of *data scientists,* who are not involved in the substantive research questions that drive the collection and structuring process.[5] In the second stage, the data or structure can be further analysed in terms, for instance, of classification. Here, various ML methods are often implemented to identify statistical patterns in the data. Lastly, steps one and two can be integrated in an automated system of data collection, structuring and classification, to produce recommendations for a certain action. Incorporating this third step is normally associated with the idea of AI systems.

Digital social science opens for a more data-driven approach whereby empirical regularities and patterns found in these new data sources may challenge existing theories and induce new theories/hypotheses to be tested. This brings tensions to the field of social science, as ML methods seek to predict outcomes based on empirical correlations observed in the data, rather than testing theory-based causal relationships (traditionally the realm of the social sciences). Digital social sciences are hence more open to taking identified empirical regularities as a starting point for research, without requiring reference to theoretically explainable causal mechanisms (Grimmer et al., 2021).

2.2 Structuring text data: an example

Digitalized information reduces any information to a specific amount and combination of bits 0 and 1 stored on a computer hard disk. Each letter in a digitalized text or the pixels in a digital photo/film can be hence defined as variables to be analysed with quantitative methods. However, the data must first be structured for the desired analytical aim using *algorithms*, i.e. rules of sequential actions to be taken in data-processing,[6] to order the information in a certain way. In automated content analysis, specific words are considered variables (columns); the units of observation (rows) are the given units of text (e.g., sentences in a book, the whole book itself, social media messages) – a process often termed *tokenization.*[7] The number in a cell hence reflects the number of times a word, or combination of words, appears in the unit of text. Words

[5] For instance, many master's degree programmes in data science offer specializations. Data from different fields are organized in different ways; and knowing the specificities is important for the work of structuring.
[6] 'Algorithm is in mathematics and data processing a complete and accurate description of the procedure for solving a calculation problem or another problem» which is contrasted with «the heuristic method where experiments, assessment and judgment along the way determine the further course of the work', translated from the Norwegian text https://snl.no/algoritme
[7] A token is the unit of analysis: thus, we can *tokenize* by word, by characters or parts of a word, or several words jointly, depending on the type of text we are interested in structuring and classifying.

and punctuation marks without major significance for the meaning may be eliminated to reduce complexity.[8]

Let us construct an example of an unstructured dataset of 10,000 sentences starting with the three following three expressions: '*The author is nice, but is unreliable'*, '*He looks nice, but detests the author*' and '*Nicer than the author*'.

Image that we seek to identify negative attitudes towards authors and possible explanations, from recorded background information of the individuals who wrote the sentence. It would be very time-consuming for a researcher to read and classify all 10,000 sentences using his/her own valuation; this would also be unreliable, as the researcher's perceptions might change in the process and induce different classification of the same text over time. An automated approach whereby a sentence will be defined as 'hateful' if it appears on a pre-defined list of words often associated with hate. This process is facilitated by the tokenization shown in Table 2.1.

Table 2.1: Appearance of words in texts, number of observations in rows and words in columns. The last column shows whether the author defines the text as hateful or not.

	Author	Looks	nice	unreliable	detest	'hateful'
1	2	0	1	1	0	0
2	1	1	1	0	1	1
3	1	0	1	0	0	0
4	1	0	0	1	1	0
1,000	0	1	0	0	1	1
10,000	0	1	1	0	1	….

Table 2.1: Appearance of words in texts, number of observations in rows and words in columns. The last column shows whether the author defines the text as hateful or not.

An ML approach differs from such automatic if–then analysis, as the starting point is the researcher's definition of the sentiment as expressed in the whole sentence rather than specific words. In supervised ML, a certain share of the sentences, e.g. 1 to 1,000, is classified, and the observed correlations between words are used to predict whether the remaining 1,001–10,000 sentences are hateful or not. If the tokens/words in the sentence correlate with those in the sentences classified by the researcher as hateful sentences, that sentence is defined as hateful, and vice-versa as non-hateful. Thus, ML is also an automated approach to classification.

Today we can construct automated classifications of entire articles, books and other texts. However, the analysis still involves quantifying inherently qualitative texts according to statistical generalizations, even if the intended meaning might differ according to such contexts as the background of the assumed reader, or form of the text (poems, scientific articles). The structure is still the same, with such quantifiable variables as columns for each defined observations as rows. However, our point here is that, similar to text, any type of digitized information can be structured into a matrix. When you purchase a product, the type, account number, data, time, outlet, etc. will be the variables (columns), and each transaction a row. Similarly, a picture will consist of a matrix of pixels, each defined as a column where the entry number indicates a specific colour and position, and each picture constitutes a row in a spreadsheet. For help in structuring such massive amounts of date, ML is very useful.

[8] Prepositions, articles and pronouns are normally considered *stop words* – words that do not contribute significantly to understanding the meaning of a text. They are often removed from the analysis.

3 Machine learning role in analysing structured datasets

3.1 Predict rather than explain

Structuring data into a matrix can be very time- and effort-consuming. *Data science* is a separate field within computer science, focused on organizing data.

With the data ordered into a structured dataset, a whole range of quantitative analysis methods becomes possible. Social scientists are interested in identifying causal relations or statistical inference between phenomena in the world. The *scientific method* involves developing a theory of how phenomenon X is related to phenomenon Y, and then testing this empirically on structured datasets to see whether this holds true in the real world. By contrast, *computer scientists* want to predict the value of the given phenomenon Y; they will then select whatever combinations of other phenomena like X will predict Y best. In the former approach, the researcher chooses the combination of explanatory variables and specific structure of the model for Y, and then runs the estimation process once. In the latter, it is the machine that experiments with various explanatory variables and model structures (within some limits set by the researchers) in repeated estimation processes, and then identifies the model that best predicts the dependent variable Y.

In theory, the same quantitative analytical method can be applied in causal analysis and predictions. However, the new methods developed within the field of ML are normally more efficient than the traditional econometric models (Mullainathan and Spiess, 2017). Importantly, as the volume of information in datasets increases, the old econometric approaches will not necessarily be able to find any solution, as processing capacity is limited, even in supercomputers. However, ML approaches are specifically designed to handle data requiring less processing capacity and should hence be able to identify solutions.

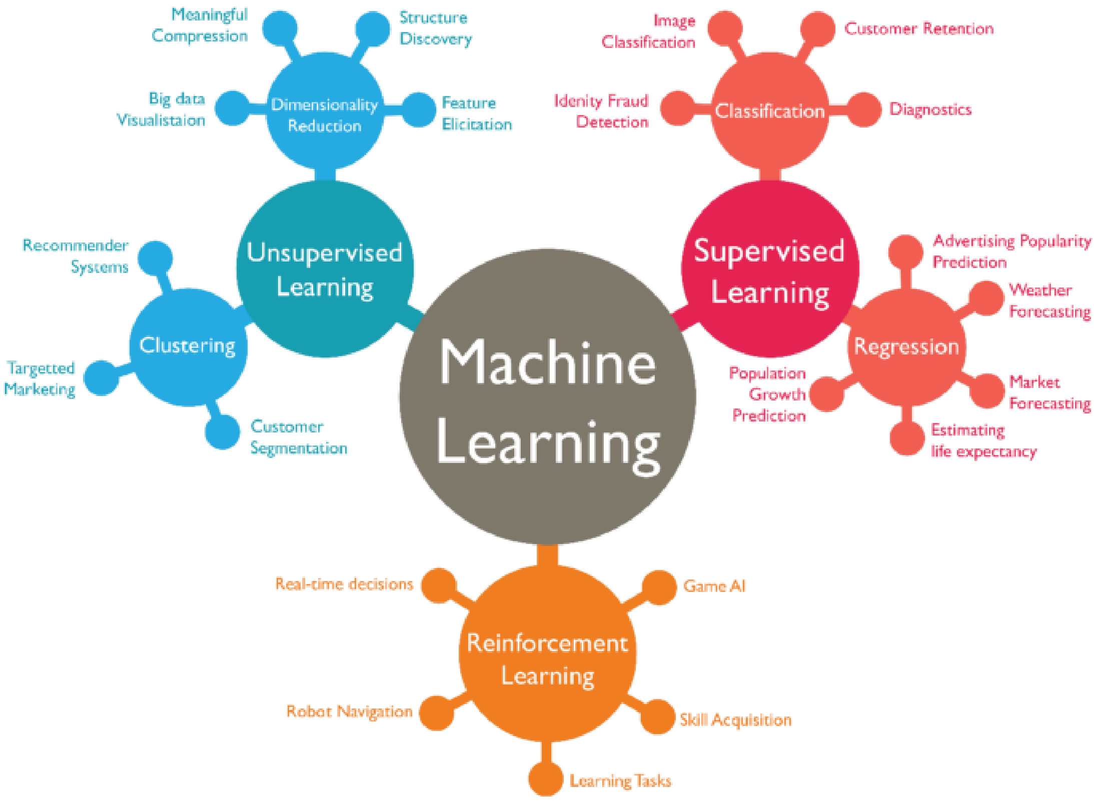
3.2 Function and types of ML

The focus on predicting outcomes rather than explaining why a given outcome occurs implies that the researcher will choose the prediction model that best fits the data. More complex models that include interaction and non-linear effects often do better, although they also make it more difficult to interpret the specific effects. The more complex the model, the higher will be the share of the variation in the dependent variable in a dataset that is explained (high *internal validity*). However, such ‘overfitted’ models are seldom able to predict information outside the dataset correctly (low *external validity*). Since the very purpose of ML in AI systems is to analyse new situations, the researcher will construct the model in way that he/she believes that will fit future situations best. The unknown future is simulated by splitting the historic dataset into three parts, for training, validating and testing.

The concepts of *learning* in Machine Learning, and *intelligence* in Artificial Intelligence, allude to human cognition and reasoning. However, as described above, these processes are purely mathematical, as the software algorithms are programmed to minimize or maximize a mathematically expressed objective function. Berryhill et al.

(2019) identify the three major categories of ML as *Supervised*, *Unsupervised*, and *Reinforced learning;* see Figure 3.1.

Figure 3.1: Applications of machine learning techniques for different tasks (Turner, 2021)



In *supervised learning* the researcher uses his/her own human cognition to define whether an observation in the datasets belongs to a certain category. The variable ‘hateful’ is introduced as the last variable in the dataset example in Table 2.1 above. The researcher has defined that the given use of words in observation 2 applies to hateful speech, unlike observations 1 and 3. Note especially that the word *detest* qualifies as ‘hateful’ in contrast to the word *unreliable*.[9] A simple approach for the researcher would then be to define all 1,000 observations (texts) as hateful or not. In the traditional social science approach, we choose an estimation method and functional form (e.g. Probit) and the explanatory variables to enter the model based on the a priori *theoretical* assumption of what causes a text to be hateful or not (e.g. reasons for including *detest* but not *unreliable*). Once we decide from theory what the universe in question looks like, the model is then estimated only once. Such method requires the variables to be normally and independently distributed for the estimated effects to reflect the true effect – a requirement that is often fulfilled. Furthermore, repeated testing of new models to identify the specification that increases the predictability implies that the researcher chooses a theory endogenously. Some academic journals hence now require that researchers deposit their theory and model specification (to be compared later with the specification actually applied) before the analysis. In that way, one can avoid the endogenous theory common (albeit deemed

[9] This subjective choice by the researcher normally reflects his/her own cognitive judgement. In this example we use ‘hateful’ and ‘not-hateful’ as a less generally agreed concept, but it is parallel to humans defining a given picture as being a ‘dog’ or a ‘not-dog’.

cheating) in social science. Such social science analysis would typically state merely that there is a statistically significant relationship between a given word and being a hateful expression for the given dataset (here, the 1,000 first observations) and emphasize that the model is not necessarily generalizable to other datasets.

However, in *supervised machine learning* the purpose is only to identify the model specification with the highest predictive power for the given dataset, and then extrapolate this relationship to define whether the other texts are hateful or not. This is a way for researchers to analyse the proportion of hateful expression in all texts they can get access to, over time, space, etc.

When the machine has been instructed to test out all sorts of specifications and then choose the one that fits the dataset best, the preferred version will typically be very complex and able to explain nearly all variation in the dependent variable. However, the model will then normally perform poorly on other data (low external validity). This problem of 'overfitting' can be solved by dividing the dataset (1,000 observations) into three portions. A given model specification is then estimated only on the first, training, portion (700 observations) and predictability then validated on the second, validation, portion (150 observations). For each model specification, the model is run several times on different portions of the dataset: e.g., another 700 are chosen for the estimation partition and therefore also other 150 for the validation partition. The overall performance of that specification will then be the average performances for a certain number of possible portions (k-fold cross-validation). In the next step, a slightly different model specification is selected, and the whole procedure is repeated, returning overall performance in the end. Finally, the chosen model is run once in the third test portion (150 observations), setting its overall accuracy.

Both social scientists and computer scientists can apply one and the same quantitative estimation method (e.g. Probit). However, the former (ideally) performs the estimation only once in order to apply the resulting model, while the latter might require millions of estimations to identify one specific model for the analysis. The second element in this ML approach involves actually applying the model to predict the dependent variable (hateful text or not) for the observations/texts outside the sample where the researcher has already defined that (first 1,000). The model has 'learned' what it takes to be hateful, and then can apply this knowledge later in defining 'hateful' texts (observations 1,001–10,000).

ML is hence used to produce variables quickly in a dataset – otherwise an insurmountable task for humans, reading through millions of texts to define them as hateful or not. The actual analysis of interest to social scientists could be, for example, to calculate the share of hateful texts over time, for certain social, cultural and geographical groups, and possibly identify causal mechanisms that trigger or deter hateful speech.

The question put in *unsupervised machine learning* would be different, and hence also the methodology. For instance, the researcher might wish to split the observations into groups with 'similar' meaning, without any ex-ante categories to classify the data. The algorithm can interpret 'similar' to mean *identical* words, where observations with many of the same words will be put in the same group. The observations are then split into different groups by a mathematical expression that maximizes the within-group correlation and minimizes the between-group correlation of the observations. Thus, unsupervised learning differs from supervised learning as the researcher applies his/her own judgement in the process of classification itself in the latter.

This approach would appear to be purely mathematical, where the ML algorithms optimize the split between groups to optimize the object function. However, the researcher will need to conduct ex-post evaluation to see whether the resulting groups have something in common that is comparable to perceived group identities in the society in question and is thus useful for social analysis. There are two different interpretations of group formation that do not resonate with human experience. Either the unsupervised ML group makes no sense in real life; or they indicate that there are social phenomena that should be identified, e.g. a basis for further investigation into why these groups do matter.

The last category is *reinforcement machine learning*. This involves a stepwise sequential simulation model where the actions taken by one actor (or by nature) in a given state (defined as a node or neuron), are responded by another actor or by nature, producing yet another state. The sequence then repeats itself in this new state, and a specific combination of sequential choices produces a result that is evaluated. An example here is the game of chess, where an action is taken by the opponent, with the outcome result of win, lose or stalemate.

However, considering all possible combinations of actions that can be taken is too complex, even with today's processors in supercomputers. The ML solution has been to predict the success of possible actions in a specific node from the experience in similar states in other games, thereby reducing the choices to be considered in the simulations. This approach has proven its worth: for example, ML chess machines easily beat traditionally programmed chess machines.

Human vs machine: Games are often used to demonstrate the superiority of machines over the human brain in information processing. It is said that the Chinese authorities started investing in AI when the reigning GO champion in Asia was beaten by a AI model. ML also discovered new strategies for chess games that human layers have adopted, e.g. Norway's own Magnus Carlsen has copied the strategy of moving the pawns on the flanks early in the game ('alpha pawn'). It may well be that in the future we will leave it to machines to test new strategies for solving challenges defined by humans before we act in line with these strategies.]

3.3 From advanced regressions to neural networks and deep learning

The first generation of ML methodologies is based mostly on known regression techniques like the Probit example above. However, by combining regression techniques in different way, we can improve the resulting predictabilities. Mullainathan and Spiess (2017, p. 90) show that ML methodologies like random forest, LASSO (least absolute shrinkage and selection operator) and ensemble (combinations of several of these methods) can predict house prices better than OLS (ordinary least squares), in both the training and the validation partitions.

Table 3.1: Performance in different algorithms in predicting house values on the same dataset, example from Mullainatnan and Spiess (2020, p.90). Note that models that fit well with the sample used in estimation (training sample) does not necessarily fit well with the set of observations not used in the estimation process (holdout sample)

	Prediction performance (R^2)		*Relative improvement over ordinary least squares by quintile of house value*				
Method	*Training sample*	*Hold-out sample*	1st	2nd	3rd	4th	5th
Ordinary least squares	47.3%	41.7% [39.7%, 43.7%]	–	–	–	–	–
Regression tree tuned by depth	39.6%	34.5% [32.6%, 36.5%]	−11.5%	10.8%	6.4%	−14.6%	−31.8%
LASSO	46.0%	43.3% [41.5%, 45.2%]	1.3%	11.9%	13.1%	10.1%	−1.9%
Random forest	85.1%	45.5% [43.6%, 47.5%]	3.5%	23.6%	27.0%	17.8%	−0.5%
Ensemble	80.4%	45.9% [44.0%, 47.9%]	4.5%	16.0%	17.9%	14.2%	7.6%

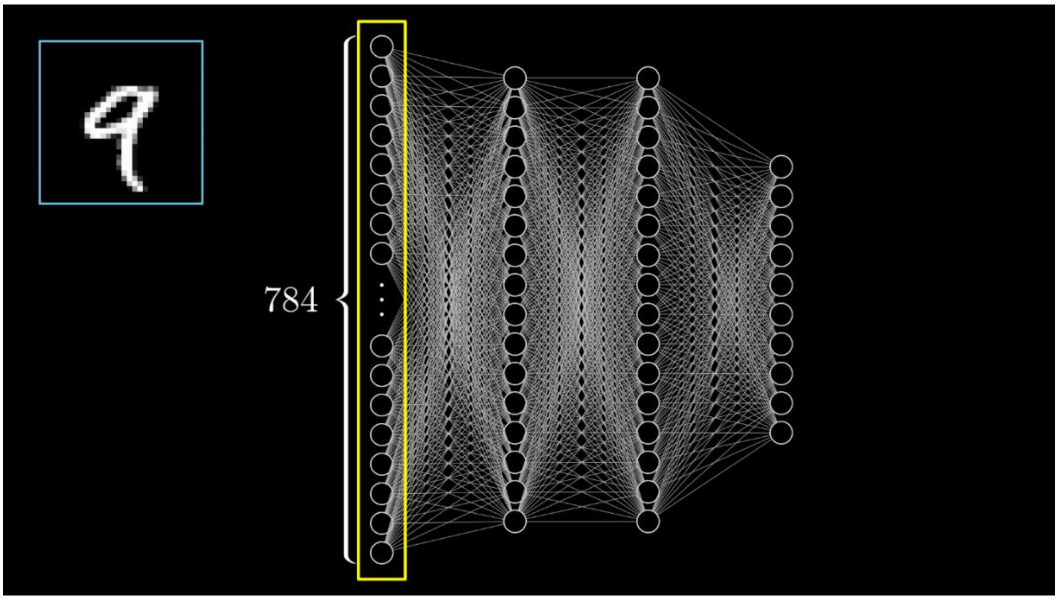
A researcher will typically choose the model that gives the highest prediction power – here the ensemble approach, as the aim with ML is predictability. The number of variables in such structured datasets is limited. However, the concept of what consists of a ‘variable’ in the analysis of pictures, sounds, etc. differs greatly. With a digital picture, each pixel can be defined as a variable, with the scale of colour tones as the realized values of that variable. Traditional ML methods are not very useful: they require considerable computation power, are slow to process and often do not even identify a solution.

This has changed dramatically with the development of the ML techniques of *neural networks* and *deep learning*.[10] The concepts again allude to human cognition, but these are also ’merely’ mathematical formulations which in the end serve to optimize an objective function. For example, a digital picture of 28 by 28 pixels will then have 784 pixels in all, where each pixel represents a variable, used to recognize single digit numbers from 1 to 9 defined as represented in the picture. See Figure 3.2.[11]

[10] The technique has long been known, but became relevant only when computer processors became strong enough to enable this technique to be applied. Basically, the strength of computer processors is set by the data storage capacity on a given space, which the computer industry has been able to double every second year over the past five decades.

[11] This example is pedagogically presented by 3Blue1Brown in https://www.youtube.com/watch?v=aircAruvnKk&t=44s

Figure 3.2: Neural network with three layers, example of number recognition by 3Blue1Brown, https://www.youtube.com/watch?v=aircAruvnKk&t=44s



Each of the 784 pixels is a variable whose value reflects the colour with value 0 for white to 1 for black, with the grey tones representing borders of the number with values in between 0 and 1. Each initial variable is hence defined as a node (neuron) with a number. In the second layer with 16 nodes, each is given a number between 0 and 1 according to a function that is the weighted sum of the 784 nodes in the first layer. Similarly, the value in each of the 16 nodes in the third layer will a weighted sum of the 16 nodes in the second layer. Finally, the value in each of the 10 nodes represents the probability of being the specific number between 0 and 9.

The machine is then programmed to identify the weights/coefficients that minimize the sum of difference between the predicted probability for each number 0–9 and the known number coded (answer book) into the picture. The whole system is hence a function where the researcher will have to estimate 13002 weights/coefficients/parameters in the function, where the initial 784 nodes in the first layer is reduced to probability of 10 numbers in the last layer in a sequential process. The concept 'neural network' is a somewhat misleading allegory to human cognition, as the impulse is unidirectional, from one neuron in a given layer to a neuron in the next layer, whereas impulses in the brain are sent in both directions between any pair of neurons.[12]

This system of equations is too complicated to be solved as a direct optimization problem. The system of equations is solved, and the parameters estimated, through numerical simulation. The machine simply tries out a random choice for all 13002 weights, and then records the payoff, i.e. the sum of differences between predicted and real outcome. These weights are then adjusted slightly in each of the coming simulations according to certain rules, to improve the payoff in the objective function. The process may also be repeated with different initial combination of weights, thereby resulting in different local payoff values. The researcher then selects the weights which yield the best result.

[12] A new generation neural network analysis includes such interaction, but introduces a time-lag in the process: signal from neuron A to B in time t, and from B to A in time t+1.

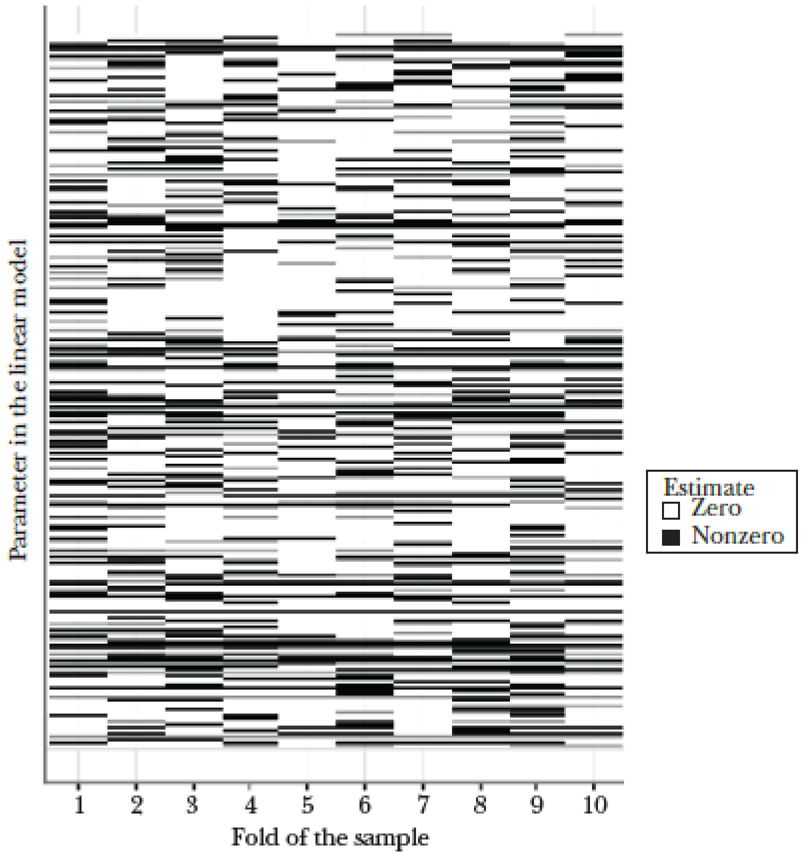
This approach has proven highly successful in picture recognition. It processes the information faster than earlier ML methodologies based on regression models, and identify with higher probability. Moreover, it can be applied for text analysis, as a text can be seen as a picture also consisting of pixels with different grey tones, and then analysed like a picture.

The same approach of neural networks can also be applied to analysis of text where words, or combinations of works, are defined as units through tokenization, as in the Twitter analysis of hateful speech example described in subchapter 2.2. above. A vector for the tokens in a given Twitter message listed in all the nodes in the first layer as in Figure 3.2 is then transformed by function in each node in the second layer. The functions in the third layer then transform the output of the function in each of the nodes in the second layer. In this example, the last layer will have only two nodes rather than nine, representing *hateful* and *not-hateful*.

The main challenge with neural networks is that we cannot really identify what causes the prediction values, and that the estimation procedures might return a different model each time. The system is hence quite unstable. However, this also applies to the simpler ML prediction models. Mullainathan and Spiess (2017, p.97) show that using LASSO methodology in different portions of the training dataset involves very different independent variables but achieves similar predictive power on housing prices (see Figure 3.3). One reason is that several of the 150 variables describing properties in the dataset are highly correlated, and are thus close substitutes and alternatives in the modelling exercise.

Geographical income distribution: Using pictures to predict economic outcomes, identify activity, environmental conditions, etc. has become recognized as an efficient way to early identify needs for intervention. An early example from urban planning is the Glaeser et al. (2016) study of how Google Street View could be used to predict the income levels of residents in urban subdistricts. With 12,000 such 360° street pictures that cover 2,439 block groups in New York City, they used ML neural network to estimate the prediction model in which predicted incomes were compared to incomes recorded in the America Community Survey data in the object function to be minimized. They used half the blocks to train the model, and then validated the model on the other half, achieving a fit of 85%. The authors argue that similar approach can be used to estimate block income levels in poor cities in developing countries, as this will require actually surveying some of them to have sufficient information to estimate the model with sufficient accuracy. Such limited surveys can have a huge potential in urban planning if combined with existing Google Street View photos as a source of data

Figure 3.3: Selected coefficients across LASSO regressions for 10 portions of the 40,000 observation dataset in Mullainathan and Spiess (2017, p.97). The 10 models differ substantially, as the non-zero explanatory variables marked in black (horizontal axis) differ between the models. It is difficult to find common features, so the ML models are less relevant, given the human need for causal interpretation.



4 Automated decisions finalize AI systems

The ultimate purpose of the ML analysis of a structured dataset is initiate optimal behaviour and choices in specific actions. The output of the ML that predicts such actions is hence only an intermediate product, to be followed by an action of some kind. If ML predicts state A, take action 1; if it predicts state B, take action 2, etc. Like humans, machines collect information (sensors vs. eyes), analyse the information to identify a state (ML vs. brain) and then act (programmed state-dependent actions vs. human decision of what action to take given the perception of the state). The machine system acts *intelligently,* although it is *artificial* and not human: this has given rise to the term *Artificial Intelligence* for the whole system of collecting, analysis and acting

As defined by the EU, AI refers to all systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – in order to achieve specific goals (EU, 2019). However, this process of collecting, analysing and acting can stop short of the last step: instead of acting, the machine can recommend a given action that humans may then chose to effectuate or not. That is still defined as AI.

Even automated actions are based on the evaluations of the outcomes from each action. As these evaluations are set by humans, it is humans that control the world through ex-ante evaluating the usefulness ('utility' in economic jargon) of different outcomes of the actions in question. Once this is done, the machine can in principle take care of the whole process.

The final step involves automated actions to create AI systems: structuring data, analysing and then taking action. The rules of action might be very simple, as discussed above; this third step of automatization is equally important as the two preceding steps, and thus merit a separate section in this report.

5 Application in social sciences

The previous section offered a brief overview of the core methods of machine learning that are commonly used by the digital social sciences. In this section, we present some examples of how these methods can be applied to areas and topics of interest to social science researchers. Here we follow Grimmer et al. (2021), in viewing ML as a tool for solving research tasks more efficiently, rather than the aim of the research activity itself.

5.1 Accessing and structuring data

One primary task of the digital social sciences is to access data. While social scientists in general are quite familiar with this step, exploring these new digital sources involves some skills not covered in most methodology courses. Engaging with *application programming interfaces* (APIs)[13] and collecting information from websites *(data mining*) are typical methods of collecting digital information that require basic programming skills. Extracting information from many documents in word or pdf format is another well-known (and tedious) way for social scientists to collect data that can be conducted more efficiently. The rapid advances in technologies such as *optical character recognition* (OCR) make it possible to transform huge amounts of digital text documents into data without the time-consuming and error-prone activity of manual typing.[14]

After accessing data, the next task is to structure them. As noted, the complexity of the structuring process will vary depending on the type of data available and the research questions involved. Extracting specific information from text (e.g., dates, names, emails) and organizing them into variables, obtaining a list of contacts of profiles from a social media platform and turning them into a relational matrix, or simply finding ways to standardize data collected from various sources or formats into a single dataset, are examples of the activities required to turn the information collected into structured data ready for analysis. Even straightforward matters like standardizing date/time or GPS coordinates in different formats can be a burdensome task.

Some of these examples will sound familiar to researchers accustomed to processing documents and text files manually. The important difference here is that the process can be automated when one is dealing with digital data. The possibility of creating routines for collection and structuring of data and automating them is one of the clear advantages of digital methods with practical application for many researchers. Moreover, by entering these routines as a code or script in a programming language, we can create pipelines[15] that are also able to collect and process new data in real time or as they become available, saving considerable time.

[13] An API is a set of protocols that enables communication between software. Many organizations are increasingly using APIs to share their data in a more structured way with researchers. One example would be that the public – researchers and others – are granted access to the content of the published webpage of the organization as a coded structured dataset and/or additional content not available on their homepages, such as statistics, background material for annual reports, etc.

[14] Some of the most popular software used by social scientists, such as STATA and Nvivo, are trying to deal with this gap by creating add-ons that allow interaction with API or have in-built OCR functions. However, their functionalities are still limited when compared to the flexibility in a programming language like R or Python.

[15] A *pipeline* is a common term used by programmers to refer to a sequence of instructions to be executed by a computer. Here, it represents the possibility of writing a code that chains the tasks of retrieving data either from an API or through scrapping a website and structuring it in a dataset.

One example to illustrate these steps concerns data from social media platforms such as Twitter. Besides being a popular platform worldwide, Twitter makes most of its data available to researchers through an encompassing and well-documented API.[16] However, to extract the relevant data from the API for a project, a researcher can use a programming language, such as Java, R or Python.[17] Once the researcher knows how to make requests to the Twitter API, retrieving data becomes a matter of defining parameters of interest for a research project (e.g. tweets containing specific terms or from specific profiles, from a specific time-period or location, etc). To make this a recurrent process, the researcher can set up an automatic update following a pre-defined interval, such as every hour, every day or week. After the data are downloaded, additional processing is necessary to structure it for analysis.[18] To identify hashtags (#) used or other profiles directly mentioned in messages, the researcher can write a few lines of code to parse the text or the variables containing the information of interest. Similarly, cleaning the text and tokenizing it for further analyses is easily done with any popular programming language. Once written, this script can be readily adjusted to gather and process new data (e.g. from a recent period or referring to a different content). The gains from economy of scale are evident.

Such skills are increasingly recognized as relevant to qualitative- as well as quantitative-oriented researchers.[19] In addition to scalability, a programming-oriented approach to data contributes to the debate of replicability of one's own research. If the researcher writes down a code, anyone may check and replicate how the data were retrieved and structured, going beyond the replicability of the model used for the analysis. Again taking Twitter as an example: any researcher who has a code containing commands for collecting, structuring and analysing the data should obtain the same results by running it from scratch, also without any knowledge of the specifications.[20] From an organizational perspective, the possibility of preserving this knowledge in a way easily accessible for different individuals is important. From an academic perspective, it helps to increase the replicability of our own research (Freese and Peterson, 2017).

5.2 Coding and Classification – Easier and more insightful analysis

Accessing and structuring new and sizeable amounts of data, of course, is not enough. Data must be classified and organized, so that one can identify patterns, interpret them and make meaningful analyses. As many social scientists know by hard-gained experience, organizing and classifying documents or interviews into categories of interest is extremely time-consuming, even with a small number of observations. There is a clear challenge when researchers with limited resources and time find themselves overwhelmed by the large quantities of data available – text in particular.

---

[16] For more information, see: https://developer.twitter.com/en/docs/twitter-api

[17] For examples of codes for requesting data from the Twitter API using various programming languages, see: https://github.com/twitterdev/Twitter-API-v2-sample-code/tree/main/Full-Archive-Search

[18] The Twitter API, for instance, provides the data in JSON format, based on Java programming language. As JSON is differently organized than traditional data frames such as the popular CSV format, they will need to be converted.

See, for example, Albris et al. (2021), who argue that programming is a useful skill for anthropologists.

[20] In the case of Twitter, there is some margin for variation due to the possibility of tweets or accounts being deleted: data retrieved on day *t* might be slightly different from data retrieved on day t+1 even if both refer to the same period (e.g. from 01.01.2020 to 01.01.2021).

Not coincidently, many digital methods focus on automating the tasks of coding and classification of data. One popular and easy way to classify the data is to use a *dictionary approach,* where the researcher uses a pre-defined list of words and expressions to classify or make his or her own. For popular analysis such as *sentiment analysis*, many established dictionaries – especially in English – indicate the overall tone (e.g. positive or negative) of a text or document. The limitation is that often the researcher must adjust sentiment interpretation in off-the-shelf dictionaries to the project of interest.

The consolidation of ML techniques brings some new tools to the kit of social scientists. In a project where the researcher knows beforehand how each case (here: text) will be categorized, the task can be considered a supervised learning problem, as discussed above. In other words, classifying texts or documents according to pre-defined categories or labels becomes a prediction exercise: to what extent does this text fall in to category A or B? By labelling a relatively small subset of the total available data,[21] we can train a model to calculate the probability of a text being in category A or B.

Use of supervised learning techniques to automate classification is rapidly becoming a popular method in digital social science. Barberá et al. (2021), for instance, provide a practical comparative guide between dictionary and supervised learning techniques for classifying large amounts of text. They conclude that supervised learning often outperforms dictionaries in terms of accuracy for sentiment or tone analysis. Moreover, supervised learning techniques can address a more diverse range of research questions, such as identifying authorship and other types of data, like images.[22]

In connection with large quantities of new data, it may also be that we lack clear theoretical expectations about how the data are organized, or that we need to update our existing categories. Take Twitter as an example again. Imagine that we collect comments about politics from new profiles, and we know nothing about them, no basis for assuming which profile will probably lean more towards conservative or progressive, or even which topics they discuss more often.

As discussed above, researchers can take advantage of unsupervised learning techniques to help them uncover patterns in new data. In that case, a cluster algorithm can be used to help to identify, for instance, common themes or topics discussed in documents, such as newspapers, parliamentary debates or social media conversations. Alternatively, as suggested by Nelson et al. (2021), these techniques can be used to validate whether the categories that appears reflect underlying features of analytical value.

This last point concerns the importance of *human* validation with use of these tools. Researchers will still need in-depth knowledge of the material to check the performance of a classification model. In many instances, the output generated by these models may also involve spurious correlations with no analytical value. Therefore, these methods cannot serve as replacements for the necessary conceptual work or the interpretation of the results. They just save some time by speeding up process that are otherwise very time-consuming.

---

[21] There is no fix minimum number of labelled units one needs to train a model. However, as a rule of thumb, the more the labelled data, higher will be the accuracy of the classification model.

[22] In the project iResponse, researchers from NILU collected pictures of wood-burning stoves from real estate ads on the popular website Finn.no and developed a supervised learning model to classify the types of stoves. The results were used to estimate air quality in specific neighbourhoods. See https://www.nilu.no/2018/06/datarevolusjon-innen-urban-atmosfaereforskning/

5.3 Social Networks

Digital social sciences are especially useful for the study of social networks. Traditionally, the study of networks and group dynamics has been extremely difficult in terms of access to data and resource-intensive procedures (Edelman et al. 2020). The flood of data from various digital sources, ranging from emails and social media, to online transactions and information from websites has had a huge impact on theories and methods for studying how individuals interact with each other in digital as well as physical spaces. Take, for example, the number of publications found in Google Scholars using the term 'social network analysis'. In 2000, we found only 819 results – whereas the average of number of results per year for the period 2016–2020 is 21,800.

It is easy to see why the study of networks has benefitted so much from the digitization of society. With easier access to data that show with whom people communicate, the content they share and the frequency of the conversation, researchers can test classical hypotheses of group formation while applying new insights into how information can spread in digital spaces. With abundant data, the field has also developed a plethora of new metrics and tools for identifying patterns of networks and community formation (Camacho et al. 2020).

For social scientists, in particular, there is considerable potential in using new sources of data in a network perspective. Hateful speech, integration of refugees and asylum-seekers, immigration flows, community organizing, political mobilization, and co-creation dynamics in policy-making are among the topics that can benefit from such approach. Sîrbu et al. (2021), for instance, note the many possibilities of using new sources of data such as social media to enable more encompassing analyses of migration flows and the factors that serve to make immigrants feel welcome in a new country, thanks to more detailed information about their social networks.

5.4 Physical geography and spatial analysis

In addition to the explosion of digital data experienced in recent decades, there has been a parallel unprecedented increase in the availability of geographic data of all types. The increase of satellite imagery, the spread of sensor and mobile phones connected to Geographical Positioning Systems (GPS) makes it possible to pinpoint locations. Consequently, the spatial analysis methods in social sciences have developed at an astounding pace (Anselin, 1999; Fotheringham and Rogerson, 2009). Visualization of spatial data, spatial statistics and econometrics have benefitted so much from the digital revolution that some authors argue for a *geographic data science* (Singleton and Arribas-Bel, 2019).

Particularly for urban settings, this deluge of geographical data has led to a further specialization under the label of *urban analytics*. Most people now live in urban areas, and most geo-referenced data will hence be from cities. Looking at this diversity, Kamrowska-Zaluska (2021) has summarized different types of urban data sources and uses of machine learning methods applied to help in illuminating questions regarding urban planning and design. However, the main challenge for this field is precisely to formulate theories that can explain the patterns found. Despite the growing amount of descriptive data showing people's movements and changes in cities landscapes, we lag behind on theoretical development that can address the why's and how's behind these data (Batty, 2019).

For social science researchers, integrating a spatial perspective in their work may seem obvious. However, the potential offered by georeferenced data (e.g. social media, mobile phones) is still open to systematic exploration. Insights from the visualization of geographical data and its systematic incorporation in statistical analysis are also examples of potential benefits to be fully unlocked. Simultaneously, the possibility of georeferencing different types of data expands the possibilities of understanding how specific neighbourhoods react to different policies, how individuals use public spaces or consume, or how local conditions affects business development.

Georeferencing of existing administrative data of companies and buildings through the connection of physical addresses with platforms such as Google Earth or Open Street Maps can be easily implemented. Ron-Ferguson et al. (2021) combine multiple administrative and georeferenced data to develop new measures of construction activity and train a supervised learning model to identify patterns of urban development in an US county as proof of concept.

Moreover, there is an untapped potential in exploring patterns of urban growth and development by analysing satellite images in combination with the rich administrative data existing in Norway, following Ogle et al. (2017) to understand sustainable urban growth in US cities and Rahnam et al. (2020) for Australia. Both studies process historical series of satellite images for different cities with aid of supervised learning algorithms to calculate populational and building density indexes, so as to evaluate to what extent these cities have adopted a more sprawled or compact pattern of growth.

Urban planning researchers are also exploring images of urban environments (e.g. Google Street View) and ML methods to map residents' perceptions of the physical space of cities where they live. Liu et al. (2017), for instance, developed a classification model of street-view pictures from Beijing to rate the quality of the infrastructure. To validate the results, they used ratings issued by people in the same locations as the pictures. Despite obtaining only average results in the model, this type of research was an innovative example of how urban planners can explore new sources of spatial data.

5.5 Construct explanatory variables for causal analysis

Traditional quantitative analysis in social science is constructed to test theories of causal effects, hence guiding the choice of explanatory variables and model formulation. In ML models, by contrast, the machine identifies the variables and model formulation that maximizes predictability; the choice is based solely on model correlations. The two approaches are hence diametrically opposed. However, programming can be useful for constructing explanatory variables for traditional testing of causal mechanisms. One example is the identification of instrument variables to deal with the two main challenges that cause biased estimates in econometric modelling. With reverse causation, changes in the dependent variable induce changes in the independent variable – the opposite of the stated research hypothesis in the econometric model, like income and social capital. It is also common to observe simultaneous changes both in the dependent and the independent variable in an econometric mode where the source of changes in both is a third 'omitted' variable that is not included in the analysis at all.

Identifying valid instrument variable will solve these estimation problems. It is demonstrated that any variable that (i) is correlated with the chosen independent variable, but (ii) uncorrelated with the omitted variables (included in the residual in the original specification of the model), will be a valid instrument for the chosen dependent variable. This means that the chosen instruments variable must only affect our dependent variable through changing the value of the independent variable of interest. We hence apply only the variation in this instrument variable that bring similar changes in the original independent variable, to measure the effect. ML can contribute to identifying a similar instrument from the dataset by predicting the independent variable, under the condition that it shall not influence our dependent variable separately. (Mullainathan and Spiess, 2017).[23]

Access to more big data also makes it possible to construct variables for phenomena that are not easily recorded in register or survey data. For example, ML models can measure, or rather predict, economic activity, identify changes in landscape using satellite images or classify companies/individuals by interpreting their written material or actions. The resulting variables from the ML predictions can be included as either independents or dependent variables in any quantitative analysis. In the end these are only variable like any other variables, independent of how they were constructed in the first place.

GPS in space planning: GPS recording applications for cellular phones has become popular in physical workout. By tracking the user's position in time, it calculates variations in speed and distance, making it easier to introduce a planned and optimized training programme. A mobility study using statistics from Strava users in Oslo demonstrates a considerable increase in people cycling and walking in the woodlands close to the city in 2020 compared to the previous year. Similarly, we find that commuting by foot or bike was higher when offices were still open, and lower when the use of home offices was enforced, giving the public authorities a way to check in real time whether the population follows the authorities' instructions or not (Barton et al, 2020)

[23] 'Natural experiments' are exogenous shocks to the independent variable of interest that are not correlated with the omitted variables. The Nobel Memorial Prize in Economics in 2021 was awarded to the three professors who developed this method, vital for estimating unbiased effects. Such studies normally begin with ex-ante knowledge about a phenomenon that can be such an exogenous shock, e.g. rain patterns.

6 Application in public administration

Social scientists investigate and identify relationships between phenomena in order to explain them. We also identify what actions are optimal in specific circumstances, e.g. state-dependent actions. An AI system simply amounts to putting the two pieces of research together. Recent governments in Norway have emphasized how efficiency in public administrative entities can be improved through the implementation of new digital tools, in communications with the public and administrative procedures (KMD 2019; KMD 2020).[24]

A first step in the digitalization process has been to organize databases and to facilitate access to these data to the public entities in their daily tasks. A natural extension was to introduce simple automated actions like sending out letters to residents if a specific task was requested in a region or at a particular time. Now we observe that public entities are moving closer to automated administrative procedures and decision-making – for example, automated allocation of resources and the first screening of applicants in a recruitment process.

The latter introduces automated judgements. The implementing agencies will need computer scientists to organize data and to formulate the ML models and AI systems. However, we hold that social scientists should be actively involved in the process of formulating these automated models – particularly those with policy implications. Social scientists have knowledge about the data of public registers and the capability of validating model outputs, knowledge that most data scientists lack.

Housing market in the Norwegian city Fredrikstad: One example is the ongoing project of projecting housing-market equilibrium until 2035 in Fredrikstad for the update of the municipal land-use and zoning plan with two components: first, to estimate housing demand based on municipal population and employment projections from Statistics Norway (also available in the Panda model) and summarize notified building projects by developers. By comparing supply and demand, the municipality can assess the need for further licensing of building permits needed to equilibrate the market. The second element of the study is to access the need for social housing. Information on income level for young adults and their parents from microdata.no will assess the number of people expected not to able to purchase their own housing but will need public assistance of some kind. Such a study can then be easily replicated in other municipalities, as the researchers already know where to collect the data and which variables are useful. There is, for example, considerable economy of scale in developing and using the same programme scripts in GIS, as the datasets from one locality may simply be substituted with the dataset with the same variables from another locality. The ease of replication provided by access to digital data is important for how social scientists approach the market for analysis in public entities.

Thus far, public entities have been understandably hesitant to introduce fully automated decision-making processes. They prefer to see the output of AI models more as another input to help human decision-makers in identifying vulnerable cases or procedural mistakes. A typical example of the former would be red-flagging of

[24] In 2020, the Norwegian government created the Directorate for Digitalization (Digdir) under the Ministry of Local Government and Regional Development, which now finances projects for introducing digitalization and artificial intelligence in public services , amongst others in the StimuLab arrangement. See https://www.digdir.no/innovasjon/dette-er-stimulab/786

individuals at risk – e.g. early warning of students expected to fail their exams – that spur an intervention that might correct the predicted negative outcome. An example of the latter is the use of algorithms in detecting potential errors in tax reports. Such construction of an AI system can be defined as research, action research or consultancy, depending on the academic perspective. Social scientists may have special competence that improves the AI system, e.g. by choosing purpose relevant dependent variables in the ML models, and detecting possibly unwanted features such as discrimination. Further, involvement in constructing such AI systems can open doors to the host institution which might request additional research related to implementation, e.g. identifying the resulting effects of the AI system implementation, highly interesting from a social science research perspective.

The concept of *digital social science* is indeed comprehensive and difficult to delimit. Here we emphasize the use of big data as well as the use of new quantitative methods to analyse them. However, social scientists can also standardize analysis that might be relevant for various public entities and municipalities.

When public entities commission more sophisticated digitalization projects to compile and structure of data from their activities, they often approach computer scientists directly, requesting ML models and AI system. However, social scientists who know the field and task at hand can better assess the relevance of models that are constructed.

Automatic refugee allocation to prevent mistrust: The Norwegian Directorate of Integration and Diversity (IMDi) commissioned a study to develop a full-fledged AI system for allocating quota refugees and asylum-seekers in 2021. The argument for automated allocation involved making better use of the historic experiences of successful allocation that what manual processing can do. This could create greater trust amongst the municipality administrations that accused IMDi of treating some municipalities more favourably by allocating them refugees with higher probability of successful integration. The AI system will have two parts. In the first, an ML model predicts the probability of success given a whole range of background information on municipalities and refugees from past allocation historically. The model then predicts the probability of success, in jobs and education at a certain time after arrival, in all the municipalities that request refugees. An algorithm then allocates all the refugees waiting, in a way that maximizes the average degree of success for all of them. Whether IMDi will apply the AI system faithfully or adjust the outcome according to the experience of people working there, remains to be seen. In the latter case, automated allocation will cease, and the ML prediction of success in pairing refugees with municipalities will be regarded more as recommendations that might be followed, or not, by the allocation officers

7 Research on the effect of digitalization on society

Beyond doubt, digitalization has increased and speeded up communication, reinforcing existing phenomena and creating new ones. However, its diverse effects require multifaceted research on how digitalization influences society and, not least, empirical research on how we can shape, through regulation, the effects of digitalization on individuals and society at large.

The different disciplines have advanced at varying speeds as regards analysing the effects of digitalization on society. Economists started early to investigate how ICT affected markets and hence the overall economy. In their literature review on *digital economics,* Goldfarb and Tucker (2019) identify causal mechanisms of digitalization leading to lower costs in (i) search of information, e.g. comparing products from all over the world (ii) replication, e.g. the marginal production cost of digital information is zero (iii) transport, e.g. near-zero for digital products (iv) tracking, e.g. facilitating logistics, also of physical products and (v) verification, e.g. the identification of both actors and product

This interest is not surprising, as digitalization has dramatically changed how markets work and hence the composition of the economy – from the introduction of auctions on Ebay, online matching services (AirBnB), skipping intermediaries to the composition of the value chain in production and the possibility of fragmenting employment in the Gig economy. Online restaurant reviews make it easier for smaller independent restaurants to be discovered by the public and hence allow for the development of *niche* establishments, to the detriment of chain restaurants (Luca, 2011). Blockchain improves certification of origin, thereby making it possible to create markets where the customer can be more certain of the sustainability of the product.

Economists are particularly interested in the effects of digitalization on job markets. The field of labour economics has been concerned about how digitalization and the rapid development of applied AI affects the labour market. It is estimated that 14% of currently existing jobs in the OECD countries will disappear, and that digitalization will change the basic character of work in another 32% of the jobs (OECD, 2019). Akerman, Gaarder and Mogstad (2015), for example, provide evidence that broadband diffusion in Norway has disproportionally benefited skilled workers. However, it is difficult to say whether we are experiencing a process of *creative destruction,* as per Schumpeter, or a structural reduction in the number and types of jobs available to workers in general. That makes it difficult to assess whether digitalization leads to higher or lower unemployment levels in the long run.

Regarding research on the consequences of regulating digital communication, economists have been especially concerned about restrictions aimed at safeguarding privacy. Kim and Wagman (2015) find that regulation on restricting the sharing of financial information increased defaults on loans during the financial crisis. Miller and Tucker (2009, 2011) showed that US healthcare privacy regulation reduced hospital adoption of electronic medical records, leading to worse health outcomes. On a more positive note in favour of privacy, Tucker (2014) shows that firm-implemented privacy controls designed to encourage consumers' perceptions of control can enhance the performance of online advertising. All regulations, almost by definition, have some kind of effect. It is difficult regulate digitalization, as the technology is developing rapidly and society is evolving with it.

Political and electoral dynamics are another example of how digitalization has changed society. The rise of social media has greatly affected how individuals consume and spread political information, with direct effects on elections. As Tucker et al. (2017) point out, the social media have become a political double-edged sword. On one hand, they have given a voice to anti-systemic groups and politicians not covered by traditional media. However, the easier dissemination of misinformation and increasing polarization (e.g. echo chambers) favoured by social media dynamics have been pointed out as negative effects of digitalization on consolidated democracies. On a more positive note, social media can be used by minorities or disorganized groups to mobilize and demand accountability from governments in less established democracies or even in authoritarian regimes. Therefore, as Boulianne (2017) points out, the political effects of social media cannot be detached from the national and local contexts where users live.

Directly related to many of the topics mentioned above, interest is also growing in understanding how algorithms and their dissemination affect choices made by individuals. Looking at recent studies about the effects of algorithms on consumer behaviour, Abrardi et al. (2021) find mixed results. While algorithms can help people to make better choices in terms of individual needs, they can also reinforce existing biases and favour dominant actors. For instance, Anderson et al. (2020) find that Spotify users who rely on algorithmic recommendations from the platform alone tend to listen to a considerably less diverse selection of music than those that do not rely on Spotify recommendations However, as Pelau et al. (2021) point out, the indiscriminate use of algorithms is heavily mediated by the social environment in which the individual is embedded. Social environments more open and optimistic about the use and effects of algorithms and AI will tend to lead individuals to adopt these new technologies less critically. Further, with the growing adoption of algorithms in the public sector, their uses and effects are becoming increasingly relevant topic for research. Is the public getting better services with AI-based solutions? or are biased implementations adopted unthinkingly by public servants? Such dynamics can summed up in the idea of *algorithmic refraction* described by Christin (2021) as a process of reconfiguration based on interactions between computational software, individuals and institutions.

The increasing importance of virtual spaces and environments offers new research areas for anthropologists and sociologists interested in identity and group formation. From pioneering works noting the importance of digital spaces, such as Castells (1996), the complexification of interactions between individuals and virtual spaces has grown exponentially. Including the virtual dimension in analysis has become increasingly important in understanding many social phenomena. Seligman and Estes (2020), for instance, summarize methodological strategies and ethical challenges related to *digital* ethnographies, and how researchers can deal with them. Platform selection, how to collect information, interaction with subjects – these are some traditional issues in qualitative research that acquire further specificities in virtual environments.

Another promising approach is to the possibility of conducting Randomized Controlled Trails (RCT) studies of policies. Introducing different regulations, 'nudging' or other interventions aimed at individuals or units like classrooms, schools, workplace, etc. are easier in a digital environment, also reducing the costs and improving the output of such research methods.

Teenagers have more friends: Longitudinal data that repeats the same questions over several decades makes it possible to identify changes in society, although not necessarily identify the causal mechanisms behind such change since different complex dynamics occur at the same time. One such large-scale data set repeated every decade in Norway is the 'Ungdata' questionnaire study of youth conducted by the Institute of Norwegian Social Research at Oslo Metropolitan University. The researchers who summarize the results point out how social media and access to digital information have changed how young people live their lives in two specific ways (Bakken et al., 2021) They now stay more at home with their families, instead of gathering in groups that 'hang out' in the street. One reason may be that they are more connected to others in social media or through gaming: that could also explain the drop in youth criminality and in alcohol consumption, and later sexual debut. However, this more withdrawn way of life has not weakened young people's sense of belonging, as they interact in even larger groups and now report having more close friends than in previous decades

8 Challenges in digital social science

8.1 Expertise needed

The core of digital social science is to make use of the large amount of unstructured data produced today in research. The first challenge is to get access to such data, as discussed in section 5.1 above. To be time-efficient and sufficiently precise, the necessary structuring normally requires fairly advanced programming skills and practice. Just as the digitalization of information into data is regarded as a specialization, the structuring of data is normally performed best by computer scientists within the field of data science.

Once the dataset is organized into a structured dataset, applying basic ML algorithms can be done by popular statistical software such as Stata, which yields a model for predicting the chosen dependent variable (Mullainathan and Spiess, 2017). Practice may make it easier to use algorithms, but the challenge is to interpret results and how the algorithm achieved them. Especially with deep-learning methods, such as artificial neural networks, calculations can seem opaque and hard to understand, even for experts in the field. For that reason, recent years has seeb the emergence of a growing field in computer science dedicated to studying algorithmic *black boxes* in to order to make them more transparent and interpretable (Samek et al., 2019).

Undoubtedly, social scientists have an important role to play in helping computer scientists to understand and interpret the results from algorithms. Contextual knowledge about the data owned by social scientists is useful to uncover potential biases due to structural inequalities embedded in the data. It might be more relevant for digital social scientists to be aware of these challenges, instead of actively engaging in trying to solve them.

Thus, the clear organizational challenge for the digital social sciences emerges: how to combine the required knowledge from computer sciences and traditional social sciences? To this, there is no single correct answer.

One obvious alternative is a more traditional division of labour between the two fields of knowledge. Computer and data scientists can be employed in social science institutes to provide technical and methodological support. [25] This alternative could be scaled up in the context of a university with the creation of a data science organization dedicated to supporting social scientists.[26]

This model, however, can be seen as a solution for a context in which there are not yet enough researchers who have been trained in social science methods but are also savvy in programming and ML methods. With the rapid dissemination of master's degree programmes focusing on data science as applied to social research, and with the inclusion of data science courses in traditional social science programmes, we can expect the coming generations of social scientists to have incorporated these methods. That also means that hiring *digital* social scientists is not far away in the future.

[25] The Peace Research Institute Oslo (PRIO) is an example, with their active policy of hiring data scientists to support researchers.

[26] The Institute for Quantitative Social Science at Harvard is an example of this model. While also conducting its own research on methodological development, its main goal is to support other departments in developing data-intensive research.

Moreover, such courses can also be taken by both junior and senior staff as part of their career development.

Despite its relatively small scale, NIBR, as an institute with social scientists, is well placed to adopt a mixed strategy. Being part of OsloMet makes it easier for NIBR researchers to engage with colleagues from the Department of Computer Sciences in developing joint projects. It also allows for the hiring of master's degree students from programmes in computer sciences to work on more programming- intensive tasks. Hiring new researchers with knowledge on digital social science methods or encouraging staff to take courses can also be a parallel strategy to ensure smoother incorporation of this type of expertise. While this strategy is more organic and controlled, it might take longer to mature, due to the limited resources NIBR has at its disposal. An alternative strategy could be a partnership between NIBR and other institutes at the Centre for Welfare and Labour Research (SVA) to scale up resources and staff into a shared organization focused on providing data-science support to all participants[27]. This could enable mobilization of more resources, but the inter-institute coordination and the definition of objectives and structure of such a new organization could take time.

8.2 Ethics and biases

8.2.1 Introduction

Access to new digital data and our increased ability to analyse them, in volume, complexity and speed, improve the possibilities of conducting high-quality, innovative social science research. Almost any possible action of humans and the state of society can now be analysed. However, this also entails huge responsibilities, as such research may have considerable impacts on others.

First, the analysis must be technically correct. Then the resulting recommendations – the actions directly resulting from the output of the analysis, must be in line with law and general ethics. In statistical analysis, only the breach of the first requirement is defined as a 'bias', whereas also the latter is defined as 'bias' in the ML and AI literature. Take statistical discrimination by gender and race. White men might do better than black women on average, something that will be picked up by the ML algorithm to predict a higher probability of success for the former than the latter, leading to preferences in automated decision-making in the final stage of the AI process. Although statistically 'true', it is still illegal to select on the basis of the individual's race, gender and age in many parts of the world. It is now also considered generally unethical to value an individual according to the behaviour of others with similar features or characteristics. This undermines all quantitative analysis which in the end are generalizations of the features, qualities and behaviour of groups.[28] However, this applies only to vulnerable groups as defined and protected by law. Generalizations for non-protected groups are legally accepted, although not necessarily deemed ethical.

The AI society limits the problem of bias in two ways. First, by restricting automated decision-making to actions which will not have effect on individuals. In that case human

[27] The Work Research Institute (AFI) has successfully applied the Norwegian research council for several large scale projects which first investigates qualitatively how AI have changed a specific sectors, e.g. policing, finance compliance and work-place organization, and then how these changes have affected public and people related to that sector, see www.algorithmic-governance.com/

[28] This also includes normal regression analysis. A significant result of a given explanatory variable, e.g. gender, may imply that people will act upon these generalizations and hence discriminate in their decision-making.

must decide whether or not to implement the machine's recommendations. Second, by disentangling the process of the ML model estimation, to identify whether protected group identity drives the outcome. Academia, authorities and NGOs are developing guidelines for 'appropriate AI' that normally at least encompass the following five main issues (Jobin et al., 2019):

- *Privacy:* prevent identification of individuals and possible leakage of sensitive information.
- *Transparency and trustworthiness:* it is a democratic right to know how decisions are made
- *Justice and fairness*: decisions must be regarded as fair to be accepted by the public.
- *Non-maleficence*: AI-models should do no harm to individuals or society, including discrimination and negative psychological, societal and economic effects.
- *Responsibility:* with integrity and respect for the law.

In the following we discuss only the first three points.

8.2.2 Access to data and individual privacy

One issue that immediately comes up when discussing the potential uses of new digital sources in research is the protection of individual rights to privacy and ownership over one's own personal data. Even though many of the new sources of information, such as social media content, are public, it is not a given that researchers can freely use their data. In the case of Norway, any project that intends to collect or use personally identifiable data must report to NSD (the Norwegian Centre for Research Data) and follow specific procedures to ensure anonymity and security.

A related challenge is the (contested) proprietary nature of many of these data sources. While some digital sources provide their data freely, many others restrict access – even to data of a public nature. For instance, public comments on Facebook – accessible also without a Facebook account – may be collected manually, but not with automated methods, according to the company's Terms of Services (ToS).[29] Although many companies have adopted the same position regarding their data, the legal validity of this restriction has been contested. For instance, Mancosu and Vegetti (2020) cite recent judicial decisions in the USA, according to which, as long a content is publicly available, any restriction on how others collect it (manually or scrapping) is invalid. Even though that decision applies only to the USA, it provides important parameters for the wider debate about accessibility to digital data.

However, concerning personal privacy, the biggest challenge is the practical impossibility of securing a complete anonymization. The possibility of linking different types of data and geographically locating those makes the identification of individuals much easier. Therefore, an important prerequisite for the further development of the digital social sciences must be a more rigorous debate about transparency and trustworthiness as to data collection and storage.

[29] The understandable reason for this restriction is to prevent other companies from monetizing the content generated by a platform or website. However, it erects extremely high barriers to researchers.

When a ML model predict probabilities, the outcome will normally be used as input in decision-making that affects human beings, either through automated decisions or as basis for people when they make decisions. AI is a very potent tool, but implementation in society will depend on acceptance by the public . People tend to accept decisions that have been made understandable: the logic and process of a decision has been presented to the public and is grasped by those affected by that decision.

'Transparent' is relative concept. ML models only maximize predictability, so it is seldom possible to relate the model to any causal mechanism by theory or experience. Most people will then not understand such mechanisms, and will therefore not perceive the model as being 'transparent' at all. However, they might still find the AI model trustworthy there is historic experience showing that AI models make better and more efficient decisions.[30]

Computer scientists will be satisfied if it is technically feasible to identify how variables affect the prediction, although the mathematical expression (and thereby real-life interpretation) can be very messy. Then they can guarantee that, for example, the model does not discriminate against protected groups, which is morally wrong and might make those responsible legally responsible for such discrimination. Even this weak definition of 'transparency' is often not fulfilled in AI, as it is hard to define what the variables in the 'black box' of ML neural networks and deep-learning algorithms reflect, e.g. by converting text to picture and thus opening for unconscious discrimination.

8.2.3 Justice and fairness

Justice and fairness are deep philosophical questions, and perceptions of one and the same outcome will differ widely. Both the notion of equality of opportunities and equality of outcomes can be direct opposites. The former will be realized in societies with highly unequal outcomes, as long as all citizens have the same opportunity to become winners, so factors like inheritance and discrimination should not apply. The moral philosopher Jon Rawls has introduced the concept 'veil of ignorance' and has shown that most people would prefer a system that secures equal outcome if they do not know ex-ante to which group/class in society they will belong, e.g. equality of outcome rather than equality of opportunities (Rawls, 2001)

Since moral ideas and perceptions differ widely, the immediate interpretation in ML and AI is that the resultant action is to be according to the law. As both ML and AI originated in the USA, perceptions of 'just and fair systems' are often reduced US law which focuses on discrimination of people that belong to a certain protected group. Only rarely has this academic field focused on any other types of perceived (in)justice and (un)fairness.

The intention of laws against discrimination in most countries is twofold: to promote equality between groups, and to prevent discrimination of individuals because of their group affiliation. However, it is possible to let group affiliation influence decisions or actions taken if : that is relevant in the case in question; it is necessary to achieve the cause at hand; and is not a disproportionate for the individual who is discriminated against.[31] The law often explicitly opens for positive discrimination for some individuals,

---

[30] One example is algorithmic stock/market trading, where speed in decision-making is essential.

[31] A useful example in the Norwegian law against discrimination Is the use of statistical generalizations to select individuals for control. It is deemed acceptable if such control is anonymous and not known even to those concerned , e.g. tax, but is seen as too intrusive if those who are controlled experience negative effect, e.g. when the police apply ethnic 'profiling' for street controls.

in order to increase representativeness within a certain sector, e.g. favouring female applicants for jobs in male-dominated professions.[32]

The main challenge within AI systems is that some seemingly group-neutral variable might be correlated with group membership. Although the group membership might not be directly recorded in the dataset, it might still affect the prediction within ML models through this correlated variable. One example would be to the use of veils by medical personnel in a hospital. Reduced ability to communicate, for physical or cultural reasons, might have a negative impact on patients, but such a variable will also pick up a possible average negative effect of being Muslim and thereby serve as an example of statistical discrimination.

If the researcher who constructs the ML model omits both group affiliation and possible proxy variables for these, predictability will be similarly reduced. The rather subjective choice of what variables to include should be guided by the foreseen applications of the results. The researcher should be very careful if the analysis entails automated decisions, e.g. sifting of applicants for a position, while inclusion might be necessary to achieve the aim of the analysis, e.g. red-flagging possible failures in school exams to effectuate early preventive interventions such as teacher support.

Moreover, the data applied in the analysis might not be very useful with regard to the problem at hand. What computer scientists define as ‘data problems’ do not necessarily mean that the data are not representative of population. One challenge for minorities is that they constitute a smaller share of the population and that the optimizing algorithms will then put less emphasis on them compared to majority, in the estimation of the ML model, e.g. face recognition where one ML model is used for all ethnic groups, despite their very different features. Some datasets, e.g. from social media, might not have some groups represented at all. Another common challenge is that the values of the variables used are influenced by stereotypes in statistical recording rather than being objective. In addition, a given dataset will report on observations made for a given time, place and group. However, the relationships identified might be valid only in that specific context and cannot be extrapolated to other observations or dataset, lacking ‘external validity’.[33]

The outcome of any data analysis will correspond to the exact question asked. If the question is not relevant for the knowledge needed, the outcome of the analysis will not be relevant either. This is interpreted as a ‘data problem’ by computer scientists even through the data are not to blame. One classical example is the process of hiring new staff in Amazon some years ago. By equating having a history of employment as a success, they found that being male was an important predictor. In fact, that probably reflected direct or indirect gender discrimination within the ICT sector – and the ML outcome simply reflected and repeated historic injustices.[34]

The ICT community was initially optimistic about the potential for AI to rather reduce than foster discrimination. Transparent mathematic optimization on objective statistics would lead to transparent and fair decisions, in contrast to human decisions where the information is filtered and interpreted by the stereotypes of individuals. Kleinsberg et al. (2018) find that an AI model would have done better than the judges in bail

[32] §11 in the Norwegian law against discrimination

[33] A common way of handling this involves continuously updating or replacing the data-set and then re-estimate the ML models, to be it as close in time, space and groups as possible.

[34] If they had tried to identify good employees instead, the same historic discrimination would probably have favoured women. If women are discriminated, they must be more dedicated or talented at any level in the company hierarchy, as they have already managed to surpass the negative stereotypes regarding women in ICT in general.

decisions, by keeping in jail more persons who came to the offence, and letting loose more persons who did not repeat the offence. Furthermore, much theoretically founded quantitative research identifies group membership as a significant explanatory variable, which is necessarily discriminatory. Once such results are made public, there is no way to prevent people from using this information to discriminate such groups, deliberately or inadvertently. Stereotypes often reflect statistical realities, so statistical discrimination becomes a convenient tool when quick, and not necessarily precise, decisions are required.

If all discriminatory selection were to be eliminated from AI systems, the outcome might be less than optimal also for the potentially discriminated. One example could be that an individual is allocated to a group where he/she will encounter discrimination by the other group members. Whether we like it or not, prejudices have real-life impacts, as they lead to discriminatory actions that affect the discriminated negatively. So the potentially discriminated may let the probability of being discriminated influence their own decision-making, for example by opting out of such situations where possible.

Another issue is that prejudices for the whole group, for example immigrants, which may be statistically true on average, cover the considerable heterogeneity within the group in question. With large datasets we often know much more about the individuals, and can thus split them into more fine-grained groups, for example immigrants by origin, age, gender, profession, and current employment status and address of residence. The error involved in treating all members of the same subgroups as ‘equals’ is less grave than doing the same for the entire group as a whole. Applying subgroup affiliation would still be discriminatory according to the law in some situations, but could be acceptable and even very useful in other situations. The predictive power of ML models can make AI systems an extremely efficient, precise tool due to the possibility of differentiating among many subgroups, while generalizing the analysis within that subgroup.

9 Conclusion

Big data constituting of all types of digital recording of human actions represents an enormous amount of data compared to what could previously be accessed for quantitative analysis. Nearly all actions and decisions made by individuals leave digital traces, hence data to be analysed, whereas register data or survey data cover only some issues and at a given point in time. *Digital social science* exploits this extraordinary rich material, enabling analyses that would otherwise be difficult to carry out with traditional data-sets.

Successful ML and AI projects often combine researchers from different fields to work efficiently. Data scientists are experts in structuring big data. Specialized computer scientists construct ML models in ways that prevent ‘bias’, while social scientists are expected to understand how society functions, putting them in a better position to decide the overall theme and purpose of the analysis that will render useful outputs.

That being said, as new generations of social scientists learn data programming in more flexible software like R and Python, more social scientists will be able to conduct all parts of the analysis themselves. On the other hand, collaboration between ICT and social science may continue as the usual solution, due to the efficiency and accuracy effect of specialization, as well as economy in scale in data processing.

References

Abrardi, Laura, Carlo Cambini & Laura Rondi (2021). Artificial intelligence, firms and consumer behavior: A survey, *Journal of Economic Surveys*. Early view: https://doi.org/10.1111/joes.12455

Akerman, Anders, Ingvil Gaarder & Magne Mogstad (2015). The skill complementarity of broadband internet, *Quarterly Journal of Economics* 130 (4): 1781–1824.

Albris, Kristoffer, Eva I. Otto, Sofie L. Astrupgaard, Emilie Munch Gregersen, Laura Skousgaard Jørgensen, Olivia Jørgensen, Clara Rosa Sandbye, & Signe Schønning (2021). A view from anthropology: Should anthropologists fear the data machines? *Big Data & Society*, 8(2): 1–7.

Anderson, Ashton, Lucas Maystre, Ian Anderson, Rishabh Mehrotra & Mounia Lalmas (2020). Algorithmic effects on the diversity of consumption on Spotify. *WWW'20: Proceedings of the Web Conference 2020*: 2155–65. https://doi.org/10.1145/3366423.3380281

Anselin, Luc (1999). The future of spatial analysis in the social sciences. *Geographic Information Sciences*, 5(2): 67–76.

Bakken, Anders, Kristin Hegna & Mira Aaboen Sletten (2021) Offline, online. Digitale ungdomsliv gjennom tre tiår, in Guro Ødegård & Willy Pedersen (eds) *Ungdommen*, NOVA, OsloMet https://press.nordicopenaccess.no/index.php/noasp/catalog/book/142

Barberá, Pablo, Amber E. Boydstun, Suzanna Linn, Ryan McMahon & Jonathan Nagler (2021). Automated text classification of news articles: A practical guide. *Political Analysis*, 29(1): 19–42.

Barton, David.N., Vegard Gundersen & Zander V. Venter (2020) *Bruk av stordata i arbeidet med å tilrettelegge for fysisk aktivitet – Kunnskapsstatus og forslag til anvendelse i Norge*, NINA report 1937, Oslo.

Batty, Michael (2019). Urban analytics defined. *Environment and Planning B: Urban Analytics and City Science*, 46(3): 403–5.

Boulianne, Shirley (2019). Revolution in the making? Social media effects across the globe. *Information, Communication & Society*, 22(1): 39–54. https://doi.org/10.1080/1369118X.2017.1353641

Camacho, David, Angel Panizo-LLedot, Gema Bello-Orgaz, Antonio Gonzalez-Pardo & Erik Cambria (2020). The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 63: 88–120.

Catalini, Christian, & Joshua S. Gans (2016). Some simple economics of the blockchain. SSRN Papers, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2874598.

Edelman, Achim, Tom Wolff, Danielle Montagne & Christopher A. Bail

(2020). Computational social science and sociology. *Annual Review of Sociology*, 46: 61–81.

EU (2019): *A definition of AI: Main capabilities and disciplines*. Independent High-Level Expert Group, European Commission, https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

Favaretto, Maddalena, Eva De Clercq, Christoph Olivier Schneble & Bernice Simone Elger (2020) What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE* 15(2): e0228987. https://doi.org/10.1371/journal.pone.0228987

Fotheringham, A. Stewart & Peter A. Rogerson (2009). *The Sage Handbook of Spatial Analysis*. London: Sage.

Freese, Jeremy & David Peterson (2017). Replication in social science. *Annual Review of Sociology*, 43: 147–65.

Glaeser Edward, Scott Duke Kominers, Michael Luca & Nikhil Naik (2018) Big data and big cities: the promises and limitations of improved measures of urban life. *Economic Inquiry,* 56 (1): 114–37

Goldfarb, Avi. & CatherineC. Tucker (2019). Digital economics. *Journal of Economic Literature* 57(1): 3–43.

Grimmer, Justin & Brandon Stewart (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3): 267–97.

Grimmer, Justin, Margaret E. Roberts & Brandon Stewart (2021). Machine learning for social science: an agnostic approach. *Annual Review of Political Science*, 24: 395–419.

Jemielniak, Dariusz (2020). *Thick Big Data – Doing Digital Social Sciences.* Oxford: Oxford University Press.

Jobin, Anna, Marcelo Ienca & Effy Vayena (2019). The global landscape of AI ethic guidelines. *Nature Machine Intelligence*, 1: 389–99.

Kamrowska-Zaluska, Dorota (2021). Impact of AI-based tools and urban big data analytics on the design and planning of cities. *Land*, 10(11), 1209.

Kim, Jin-Hyuk & Liad Wagman (2013), Screening incentives and privacy protection in financial markets: a theoretical and empirical analysis https://ssrn.com/abstract=2317942 or http://dx.doi.org/10.2139/ssrn.2317942

Liu, Lun, Elisabete A. Silva, Chunyang Wu & Hui Wang (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems*, 65: 113–25.

Mancosu, Moreno & Federico Vegetti (2020). What you can scrape and what is right to scrape: A proposal for tool to collect public Facebook data. *Social Media + Society*, 6(3): 1–11.

Miller, Amalia R. & Catherine E. Tucker (2011) Can health care information technology save babies? *Journal of Political Economy,* 119 (2): 289–324-

Mullainathan, Sendil & Jann Spiess (2017) Machine learning: An applied econometric approach, *Journal of Economic Perspectives*, 31 (2): 87–106.

Nelson, Laura K., Derek Burk, Marcel Knudsen & Leslie McCall (2021). The future of coding: a comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods & Research*, 50(1): 202–37.

Ogle, Jared, Donna Delparte & Hannah Sanger (2017). Quantifying the sustainability of urban growth and form through time: An algorithmic analysis of a city's development. *Applied Geography*, 88: 1–14.

Pelau, Corina., Irina. Ene & Mihai-Ionut Pop (2021) The impact of Artificial Intelligence on consumers' identity and human skills. *Amfiteatru Economic,* 23(56), 33–45.

Rahnama, Mohammad R., Ray Wyatt & Lia Shaddel (2020). A spatial-temporal analysis of urban growth in Melbourne: Were local government areas moving toward compact or sprawl from 2001–2016? *Applied Geography*, 124, 102318

Rawls, John. (2001) , *Justice as Fairness: A Restatement*, Cambridge, Massachusetts: Belknap Press.

Ron-Ferguson, Nathan, Jae Teuk Chin & Youngsang Kwon (2021). Leveraging machine learning to understand urban change with net construction. *Landscape and Urban Planning*, 216: 104239. https://doi.org/10.1016/j.landurbplan.2021.104239

Samek, Wojciech, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen & Klaus-Robert Muller (2019). *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Cham: Springer.

Singleton, Alex & Daniel Arribas-Bel (2019). Geographic data science. *Geographical Analysis*, 53(1): 61–75.

Sîrbu, Alina, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini (...) & Rajesh Sharma (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11: 341–60.

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts & Pablo Barberá (2017). From liberation to turmoil: Social media and democracy. *Journal of Democracy*, 28(4): 46–59. https://doi.org/10.1353/jod.2017.0064

Turner, Louise (2021). Machine Learning: A Primer. https://medium.com/@lizziedotdev/lets-talk-about-machine-learning-ddca914e9dd1