MAJOR ETHICAL ISSUES IN THE DIGITAL ERA

A co-word network analysis to monitor the evolution of biases and ethical issues behind algorithms & AI during the last 10 years

Ivan Grazioli

Master's degree in Strategic Innovation Management

University of Sussex, Brighton, UK

September 2020

UNIVERSITY OF SUSSEX | BUSINESS SCHOOL

Abstract

This paper reports the findings of a scientometric co-word analysis designed to uncover main topics around AI and Ethics have been at the centre of the debate during the last 10 years. It presents an introduction and a background study related to AI and ethical issues from academic literature, followed by SNA concepts and methodology description. Biennial networks visualisations and aggregated ones have an important role in terms of results and discussion since they let the reader perceive the actual increase of number of publications and the main terms and topics encountered. Conclusions and recommendations would serve as a motivation to invest in improving the design of an ethical and legislative system in relation to the digital governance, and in increasing the interest of young people in the world of politics, in order to have future leaders aware of the problems analysed in this paper, and who have the mission to build a world a better place than they have found.

Table of Contents

List of Abbreviations

3D	3 DIMENSIONAL
AI	ARTIFICIAL INTELLIGENCE
CEO	CHIEF EXECUTIVE OFFICER
DB	DATABASE
EDA	EXPLORATORY DATA ANALYSIS
ELSI	ETHICAL, LEGAL, AND SOCIAL ISSUES
ERP	ENTERPRISE RESOURCE PLANNING
GDPR	GENERAL DATA PROTECTION REGULATION
GML	GEOGRAPHY MARKUP LANGUAGE
HR	HUMAN RESOURCES
IoT	INTERNET OF THINGS
KPI	KEY PERFORMANCE INDICATOR
ML	MACHINE LEARNING
OECD	ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
SNA	SOCIAL NETWORK ANALYSIS
WoS	WEB OF SCIENCE
XML	EXTENSIBLE MARKUP LANGUAGE
YOY	YEAR OVER YEAR

List of figures

Acknowledgments

I would like to thank the following people and institutions, without whom I would not have been able to complete this dissertation, and without whom I would not have made it through my master's degree.

Firstly, thanks to the University of Sussex *Science Policy Research Unit*, especially to my supervisor, Professor Simone Vannuccini, whose insight and knowledge into the Artificial Intelligence matter steered me through this research.

A special thanks to Janina, my landlady, not only for her proofreading services, but also for having hosted me in the amazing city of Brighton. Thanks to Melina, always ready to bear with me and to support me during small daily discussion.

Thanks to my Strategic Innovation Management classmates, coming from all over the world, for having taught me how diverse and beautiful our planet is. You are the real take away I gained after having attended this course. Our connection will last during our entire lifetime.

Biggest thanks to all my friends from Bergamo, especially to Chiara, Dario, Francesco and Giulia. For all the support you have shown me through these months.

My ultimate thanks go my anchor, the incredible woman from whom I received my life's blood, Lucia. My mother never gave me any idea that I couldn't do whatever I wanted to do or be whomever I wanted to be. She filled our house with love and fun and food (often junk food). I thank her for all her hard sacrifices and for showing me how much effort in life counts to get results. Her most important teaching. No matter how unfortunate you may be, the effort always pays off.

1. Introduction

When it comes to Artificial Intelligence and to the impacts it has (and could have) on human beings, the positions are very different. On the one hand we find those who say that AI is incredible, it will make us go beyond our biological limits, we can live forever, become immortal. We could do mind uploading, taking all the information from our brains and transfer it to a machine, becoming so very intelligent and very powerful, eternal. On the other hand, there are those who say that AI will lead us to have more powerful and refined machines, able to define and create the rules of their design, and so becoming more and more powerful and more and more refined that will eventually make the human species superfluous. They will imprison us, they will exterminate us or they will enslave us. Here, the transhumanism and the singularity explained.

All of this is science fiction. Nothing to do with research and nothing to do with the reality in front of us. What Artificial Intelligence has brought us today, in fact, was, in short, the ability to perform very complex tasks in a much faster and more efficient way than us as human beings. Obviously doing so without having a true intelligence at the base: no intuition, no creativity, no ideas, no empathy, no awareness or consciousness.

From an ethical point of view, having machines that perform tasks without having the characteristics listed above can become a problem. In this paper, these types of ethical issues will be analysed excluding a priori everything concerning the medical and biological application of AI, trying to define a comprehensive and temporal vision, explaining how they have evolved over the last 10 years. All aspects related to privacy, ethnicity and gender issues will be taken into consideration.

Thanks to Social Network Analysis (SNA) methodology and, specifically, a scientometric time-range co-word analysis, it will be possible to note which problems and topics related to these issues have become relevant over time and how, during the last five years, there has been a boom in publications related to the topic. To converge to a solution for these problems in our modern society, this paper relies on Luciano Floridi's *naïve ideas* (Floridi, 2020) suggesting the design of a good and informed governance as preventive and necessary measures.

2. Literature Review

a. Background

"Scientia potentia est". The concept *knowledge is power* finds its Latin roots a few years before the born of the Enlightenment period. The phrase is commonly attributed to Sir Francis Bacon in *Meditationes Sacrae* (1597). However, a similar thought can be found in the Bible's *Book of Proverbs*, Old Testament, written back in 700 BC. It states: "*A wise man is strong, a man of knowledge increaseth strength*" (Book of Proverbs 24:5).

To better interpret the role of knowledge in our modern society, in 1989, professor Ackoff designed a hierarchical pyramid, which finds its base in data (Figure 1). Wisdom, as highlighted in the figure, can't be generated by data directly. It needs a set of models and processes to collect, manipulate, clean and analyse data in order to extract information, which, once put together, could generate knowledge.

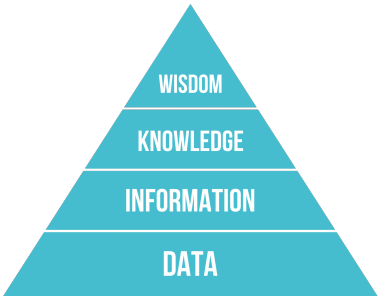

***Figure 1**. DIKW hierarchy as a pyramid (Ackoff, 1989)*

In 2004, another schema was developed to allow us to view data, information and knowledge as distinct kinds of **economic goods**, each possessing a specific type of utility. An example of these goods can be telephones or books as paradigmatically data goods; specialized newsletters exemplify information goods; and brain surgery can be thought of as a knowledge good (Boisot & Canals, 2004). Basically, information is an *"extraction from data that, by modifying the relevant probability distributions, has a capacity to perform useful work on an agent's knowledge base"*. Their essential relationships are represented in Figure 2.

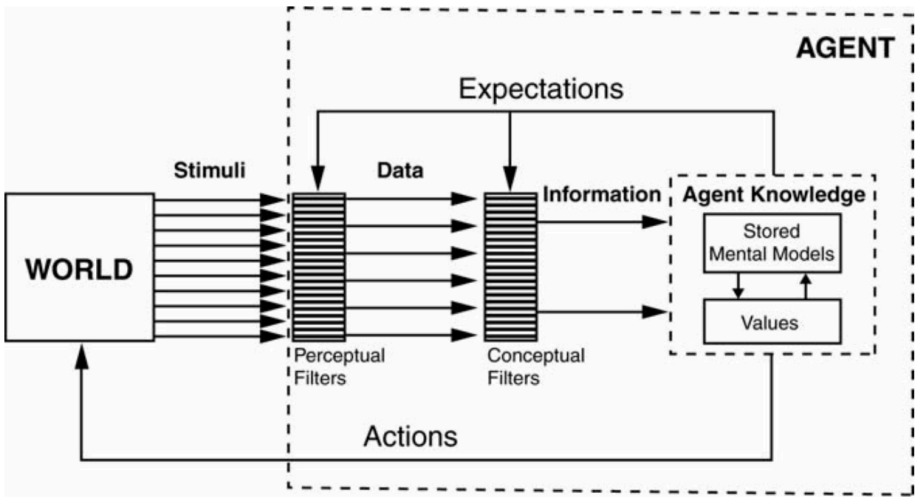

***Figure 2**. Data, Information and Knowledge in the world.*

Data became more and more important during the last decades. The Economist said in fact that "The world's most valuable resource is no longer oil, but data" (The Economist, 2017). It can be both generated by people or devices themselves. Below, Figure 2 and 3, describe the growth of the number of individuals connected to the internet and the amount of connected devices in the world. In 2014, the number of active mobile devices crossed over the number of human beings, reaching a total of 7.19 billion smartphones. Today, according to GSMA Intelligence data, there are more than 10.3 billion cellular connected devices (GSMA Intelligence, 2020).

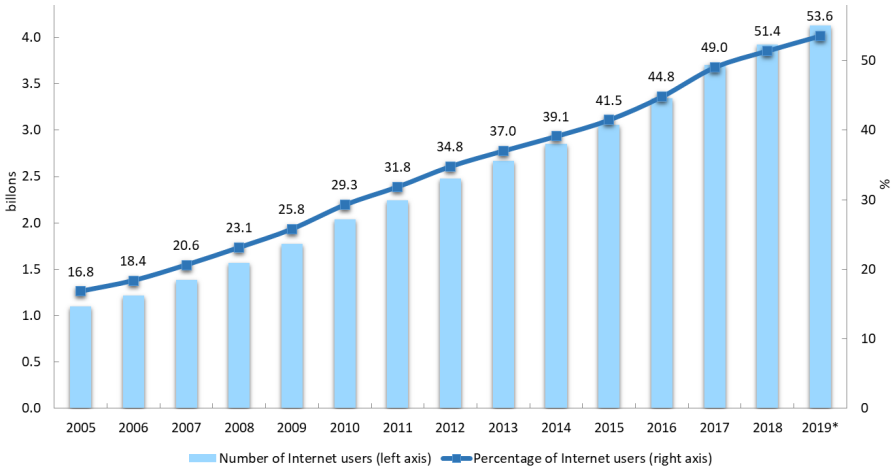

***Figure 3**. Individuals using the Internet, 2005-2019. Source: ITU*

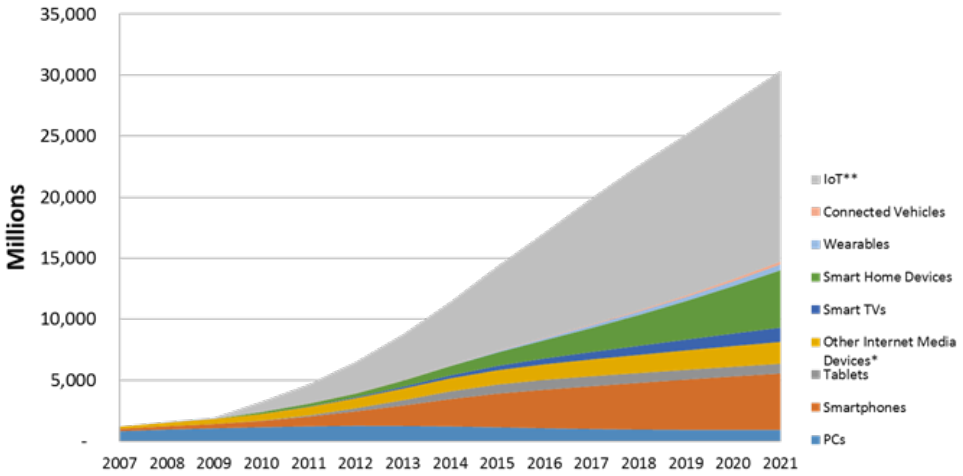

***Figure 4**. Global connected devices, 2007-2021. Source: Strategy Analytics research services, 2017*

As anticipated by the Ackoff's pyramid, and stressed by David McCandless, data can be identified not as the new oil, but as the new *soil* (McCandless, 2010).

The huge amount of data generated every day by humans and machines is stored in a disparate way. This can be compared to a brainstorming exercise (divergence), where bits of information are generated and stored in different systems, different locations and different formats across the globe. For example, a hospital in Italy stores data related to patient A and another hospital in London stores data related to patient B. Unfortunately nowadays it's impossible to immediately compare or cross-analyse data related to the two

patients due to structural, legal, geopolitical and many other reasons. Here is where entropy (the Shannon's one) comes in and it can be identified as a measure of uncertainty (Kwiatkowski, 2018). This term, in Information Theory, is related to the average level of "surprise" or "uncertainty" of a random variable. So, without proper action to analyse, study and interpret those bits of information (convergence), they would remain pretty useless by themselves.

Thanks to the advent of the digital era and the exponential growth of computing capacity of modern machines, in which the number of transistors in microprocessor circuit double approximately every two years (Moore, 1965) data analysis and, recently, artificial intelligence, have found space to transform complex procedures into something that can be done much faster or even automatically. Let's see how.

b. The evolution of AI, from the Turing test to the World Envelopment

Everyone seems to be talking about Artificial Intelligence (AI) these days, but it's good to remember that it is not something new. Back in 1950, Alan Turing proposed the Turing test and that same year Isaac Asimov published the three laws of robotics. The term AI was coined firstly by John McCarthy in 1955 and in 1956 became a new science field.

60 years of Artificial Intelligence evolution brought the field to a broad family. Below, Figure 4 describes AI different typologies, clustered in 3 main paradigms: Symbolic, Statistical and Sub-symbolic. Normally, when someone thinks about AI, there is an immediate connection to robotics, imaging a Star Trek cyborg relating to humans and able to think intelligently. This is not the case. Massively most of AI is not embodied. The real revolutions that made the difference in the past few years are Deep Learning and Generative Adversarial Networks. They fall under the typology of Neural Networks within the Machine Learning cluster. But how do they work? Reducing the complexity of their actual behaviours, we can say that their main task is to map data and information. They take massive quantities of data as input and provide information as output. An example that can be seen in everyday life is the recognition of images. A famous storage platform, for example, allows the search by keywords of images taken by the user, recognizing the subjects (and often also the text) of the photograph.

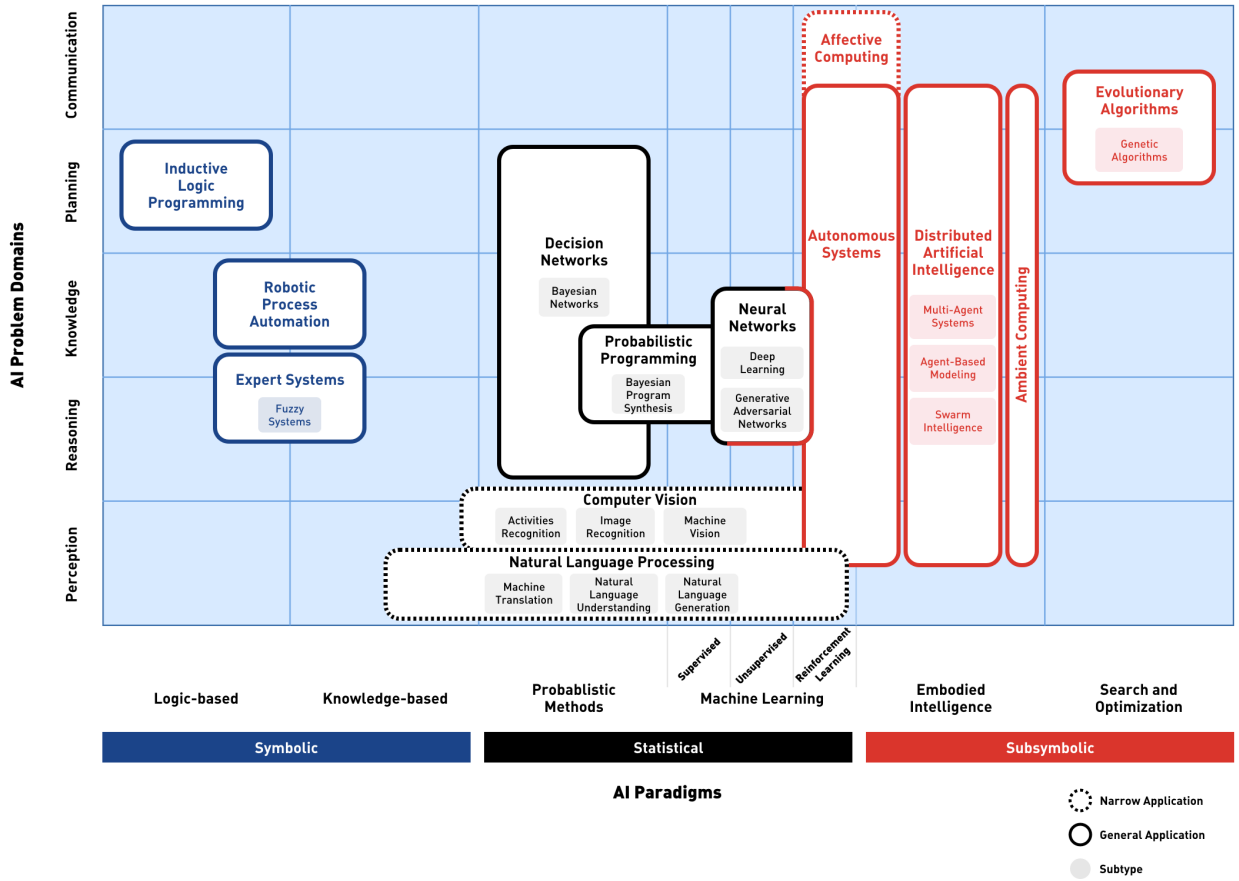

***Figure 5**. AI Knowledge Map,* (Corea, 2018)

Having identified AI as a family, we should focus on its functionalities. As Luciano Floridi argues, "*AI is a divorce between the ability to solve successfully a problem and the need to be intelligent at all in doing so*". In fact, focusing on the machines' cognitive ability of doing things in an intelligent way (and instinctively) is currently in its infancy, since today we are not yet making great progress. On the other hand, when it comes to engineering, automation and machine computation are like a magic spell. How and when did AI get so much power to complete a certain task? One of the explanations of the success of an engineering approach of AI is called an envelope. Basically, to let machines do things better and faster than humans, we decided to build a 3D space within which a particular engineered artefact is successful. So, with the advent of those advanced AI and machines, we are not so much adapting the technology to our world, but we're changing and building new environments letting them perform their tasks at lower costs, reorganising the world around the ability of a machine to do certain things (Floridi, 2019).

Having the AI role clear in mind in our modern world, the attention should be moving on to what are the risks and the opportunities related to its usage and which major ethical and political issues can arise with the good or the bad uses of the new digital technologies and their governance? The next three subchapters will investigate these questions.

c. AI Risks and Opportunities

As per human behaviour, AI algorithms are similarly driven by **nature** and **nurture** (Hosanagar, 2019). Nature is actually the logic that a programmer codes in the algorithm or the set of rules that is follows. Nurture, instead, is the data which programmers rely on (or rather on what they are trained on – e.g. Machine Learning). As Hosanagar points out, nature or nurture, either way, the parent is always going to complain (in our case the data scientist).

So, relying on this *parent* point of view, here below are described the main risks of AI. They are clustered around 5 main **risks** that can affect a system (or a society) in general. Social Risks include:

- Labour displacements
- AI Oligopolies
- Surveillance and control
- Discrimination
- AI is underused (due to opportunity cost)

All of those risks have some common characteristics. AI, when misused or overused, can, in fact, have the following negative effects (Floridi, 2019):

- Devalue human skills
- Removing human responsibility
- Reducing human control
- Eroding human self-determination

In terms of **opportunities**, AI incredible precision, accuracy and speed in completing tasks could take place of humans in repetitive and tedious duties in different environments, even dangerous ones, since it won't be affected by hostile environments. As anticipated above, AI is also a great tool that can help us to convert information into knowledge, being able to analyse an extensive volume of data. Moreover, during the last few years, great improvements are being made for images recognition. There is a very wide range of applications areas, such as medicine (recognition of malformations from body scans) or automated industry (recognition of mass-produced defects).

From an economic point of view, those opportunities have solid profit motives since performances optimisation brings to a revenue growth, somehow related to the western capitalism we know.

However, nowadays we are facing many issues as a species, climate change being one such issue. Here AI could help us in developing and acquiring coordination, to tackle those big issues together.

d. From the Governance to the Design

During the last few years, many frameworks, guidelines and list of AI principles have been developed by many governmental and non-governmental organisations (e.g. the OECD). Others include the European Commission's *Ethics guidelines for trustworthy AI*, developed by the High-Level Expert Group (European Commission, 2019) and the *Beijing AI Principles* backed by the Ministry of Science and Technology and

drafted by the Beijing Academy of AI (Ministry of Science and Technology, 2019). So, we can see clearly how important the governance and the regulations behind AI are, in order to be deemed trustworthy.

If we dig deeper into the academic world, a network of publications has been developed to identify which are the main topics related to governance, the public sector and AI. A Web of Science list of academic articles has been gathered and filtered over the last 5 years (from 2015 to 2019) with searches of the main keywords related to "AI", "Governments" and "Regulation" (and their derivates). The clusters automatically generated can be interpreted as follows:

- Green: AI Issues
- Blue: AI Law
- Purple: AI Applications
- Yellow: AI Environment
- Red: AI Infrastructure.

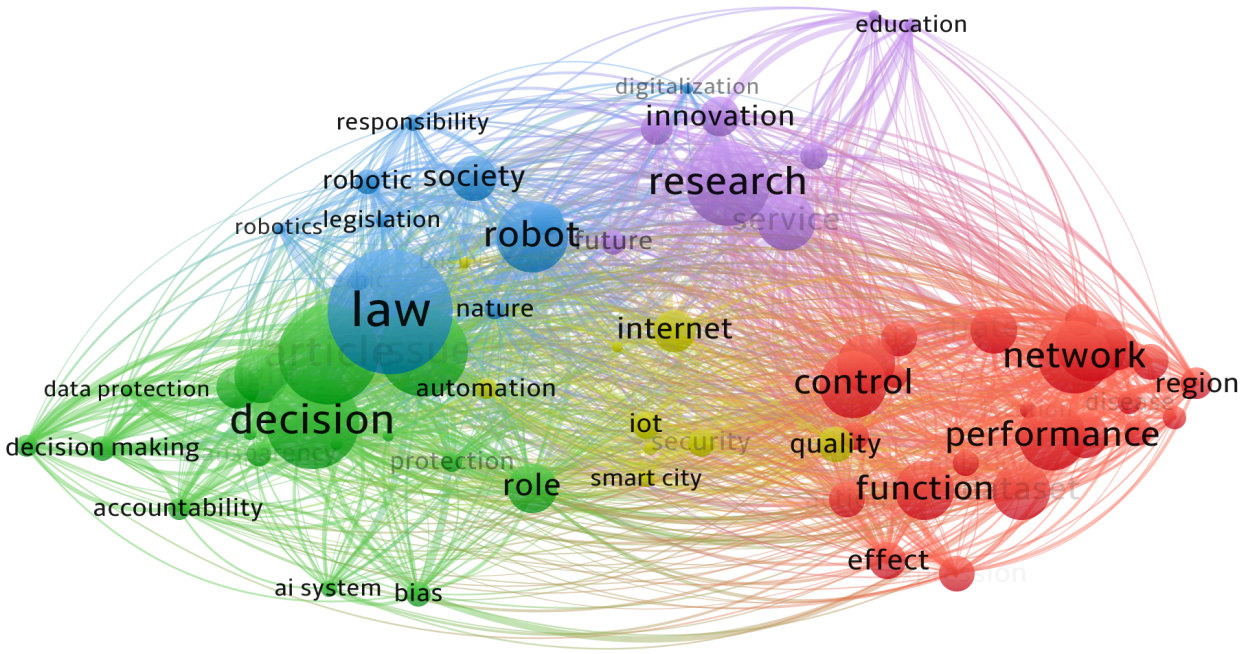

***Figure 6**. VOSviewer network of a co-word analysis of publications on AI, Regulations and Governance between 2015 and 2019*

From a geographical and an organisational point of view, the following network describes the connection between entities publishing papers related to the topic in question. Nodes' size represents the number of documents published, nodes' colour represents the organisations' country and the edges identify the citation exchange between them.

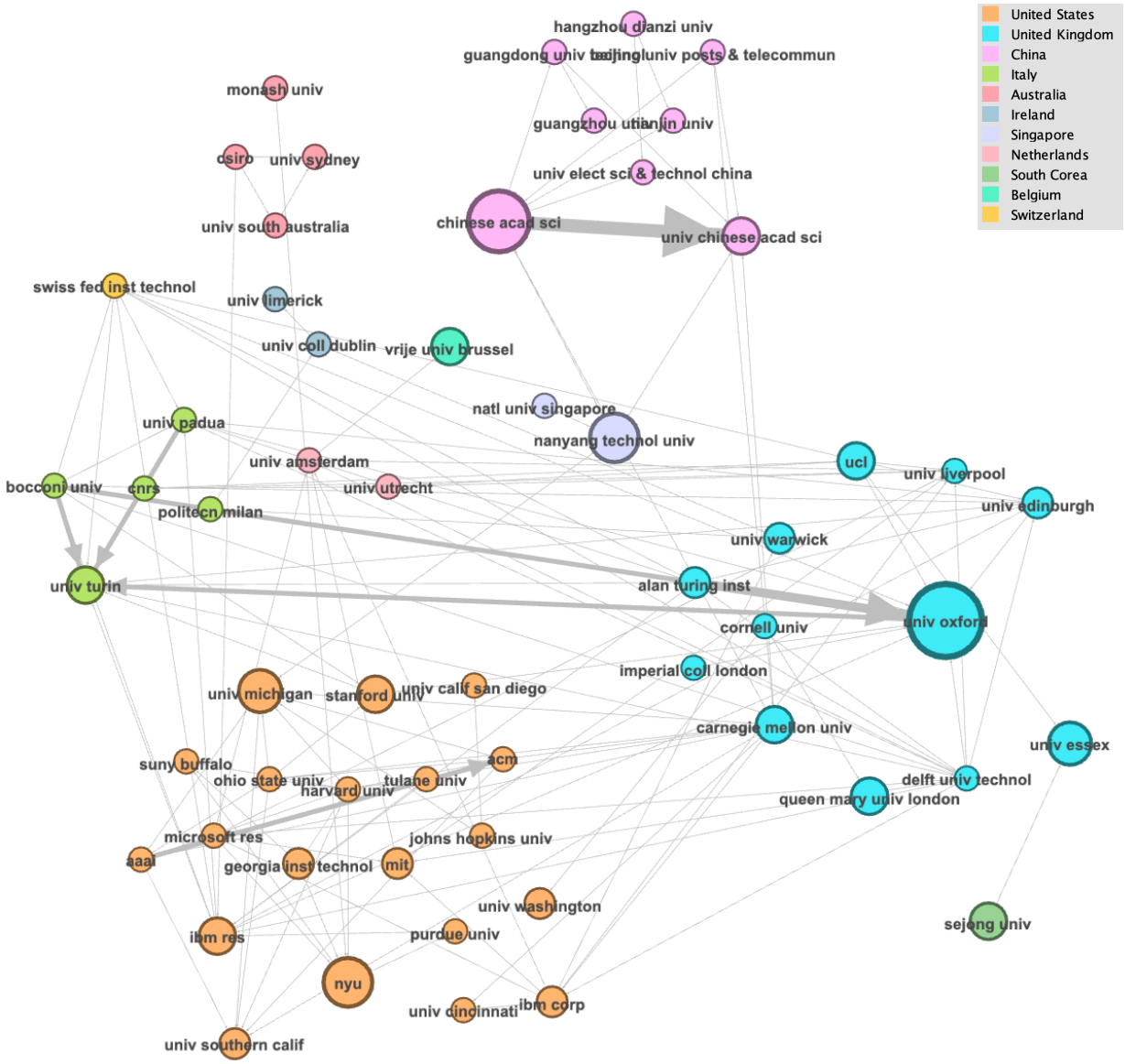

***Figure 7**. Gephi visualisation of Organisations' publications on AI, Regulations and Governance between 2015 and 2019*

From the network we can identify that during the last 5 years 11 countries and 58 organisations published at least 2 papers cited at least 1 time by other institutions. United States, United Kingdom and China are countries where there are more organisations and where more content has been produced.

The effort spent by those actors in studying and researching on this topic will drive new regulations in order to design better policies related to the AI development and infrastructure. This is also because, as Floridi says, *"governance is a matter of design"*. And governments, above all have, have the hard duty to design what kind of society we want.

e. Main ethical challenges

Recent literature highlights how an ethical approach could bring a *"dual advantage"* (Floridi, et al., 2018). On one side, ethics enables organisations to take advantage of the social value that AI enables. On the other side, ethics enables organisations to anticipate and avoid or at least minimise costly mistakes. So, searching

in the literature recommendations—to assess, to develop, to incentivise, and to support good AI- can be found. But what happens when they're not respected or at least observed? Well, different risks have been highlighted in the previous chapter and to tackle them, we rely on the above referenced *AI4People* Ethical Framework, and specifically its point #14, below, highlighting 2 important topics related to diversity ethical issues.

> *14. […] Ethics should help seize opportunities and cope with challenges, not only describe them. It is essential in this respect that diversity infuses the design and development of AI, in terms of* ***gender****, class,* ***ethnicity****, discipline and other pertinent dimensions, in order to increase inclusivity, toleration, and the richness of ideas and perspectives.*

If we consider the people at the centre of this study, another important area should be taken under consideration here: **privacy**. Here, a wide literature has been produced (at least in Europe) since the advent of the GDPR (General Data Protection Regulation). Also new consultancy firms and start-ups have been created in order to help customers in making AI compliant with GDPR.

So, the three main ethical issues falling under the diversity and the fairness perspective are: gender, ethnicity and privacy. Nowadays the debate is pretty heated both at academic and at newspaper level. Below a few examples.

Gender

A recent example from PwC (PwC, 2017) American, European, and Asian international organizations continue to outsource about 40% of human resource functions to the algorithms. The program was based on data extrapolated from the resumes of people hired in the last ten years - almost exclusively men - and consequently automatically discarded women.

Another example can be easily identified on Google Image search. In fact, looking for the keyword "CEO", all the resulting image in the first page includes men only.

Ethnicity

"Broken data means you get Broken Barack Obama". This example finds its root in AI image recognition algorithms. Shortly after publishing the research, some Twitter users started probing the model for biases. And they found some unfortunate results. The image below is the result of a blurred Barack Obama transformed in a white man.

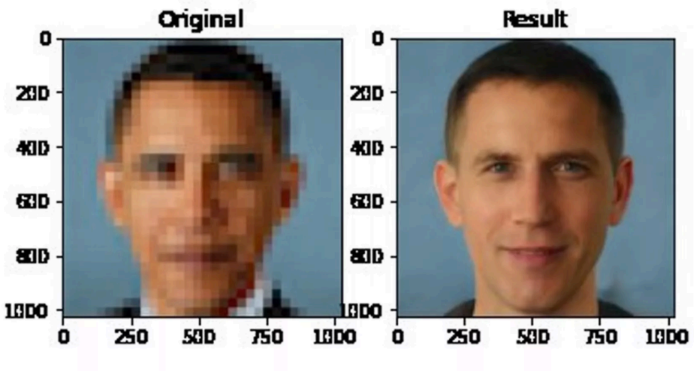

Figure 8*. Upscaling Barack Obama into a white man* (Sütterlin, 2020)

Privacy

This next example has probably also happened to you and doesn't need a reference. Facebook, Amazon and other companies AI algorithms keep collecting our personal information related specifically to our navigation data and cookies. We can be considered all as Data Subjects. Those companies collect all the data they can in order to create profiles and suggest and propose a new product we could be interested in.

As mentioned in a paper developed by AINOW Institute at the New York University in 2019, one of the main recommendations to the AI industry indicates a urgent need to make significant structural and technological changes in order to address systemic discrimination in terms of racism, misogyny, and lack of diversity (AI Now Institute, 2019).

3. Methodology

a. Data

The scientometric co-word Network Analysis presented in this paper relies on a dataset of journal articles downloaded from *Web Of Science* (WOS), an online subscription-based scientific citation indexing service, originally developed by the Institute for Scientific Information, now owned by Clarivate Analytics, which provides a comprehensive search engine for academic publications.

The data cover a 10-years' time range set of papers, from 2010 to 2019. The extracted records fall under the academic journal typology. Below is the query developed to gather all the publications which contain meaningful keywords for the analysis.

```
(AB = (ethic* AND AI
        AND (gender OR race OR ethnicity OR privacy OR divers*))
OR AB = (ethic* AND "Artificial Intelligence"
        AND (gender OR race OR ethnicity OR privacy OR divers*))
OR AB = (ethic* AND algorithm*
        AND (gender OR race OR ethnicity OR privacy OR divers*))
OR AB = (bias* AND algorithm*
        AND (gender OR race OR ethnicity OR privacy OR divers*))
OR AB = (bias* AND AI
        AND (gender OR race OR ethnicity OR privacy OR divers*))
OR AB = (bias* AND "Artificial Intelligence"
        AND (gender OR race OR ethnicity OR privacy OR divers*)))
NOT AB = (medic* OR patient* OR radiolo* OR clinic* OR biolog*)
```

The query's keywords have been searched within the abstract (AB) of each article contained in the WOS database. The first part of each line aims at selecting all the papers that have at least a couple of words related to Ethics, Artificial Intelligence, Algorithms or Bias. A further filter has been applied to refine the results, focusing on at least one diversity or privacy issue (*gender OR race OR ethnicity OR privacy OR divers* OR racial*). The reason why this has been added is determined by the need of tracking those kinds of ethical issues, bypassing other topics not relevant for the analysis, which would have blurred and worsened the dataset with irrelevant words. The last line excludes all the papers related to medicine which contain words like medic, patient, radiology, clinic and biology.

As a first insight, we can see that the query result contains a set of **452 articles**. From a publication year and a time range perspective, a regular increment of publications every two years represents the growing concern related to the topics analysed (Figure 8). On a country basis, the United States, China and England can be identified as top locations where those studies have been developed (Figure 9). On average, each paper extracted includes 45 cited references, has been cited 20 times and is 15 pages long. The organisations with the highest number of published articles are the *Chinese Academy of Science* (13), *Stanford University* (8), *University of Florida* (7) and *University of Oxford* (7).

To properly create five biennial networks, the 452 records will be downloaded separately for each specific biennium, in order to be imported in the tool used for the network creation.

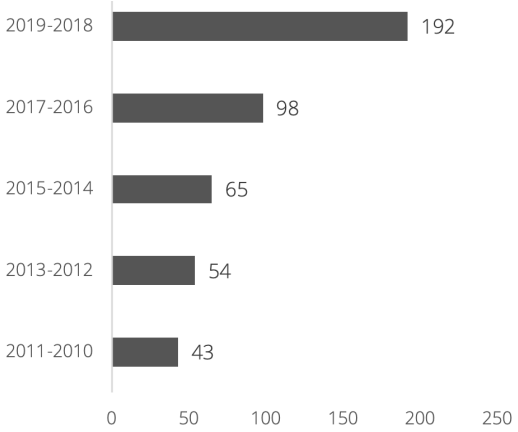

***Figure 9**. Number of publications per biennium*

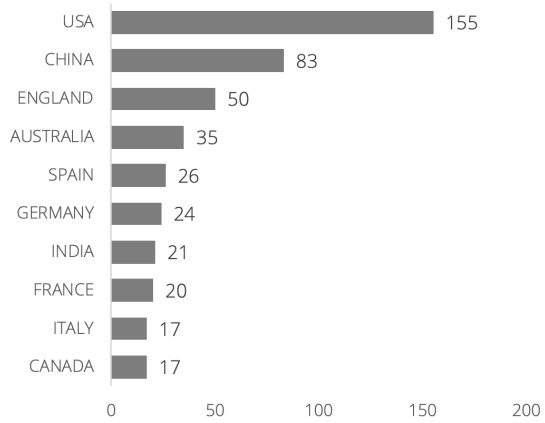

***Figure 10**. Number of publications count per country*

b. Tools

The tools used to create the five co-word-analysis networks are *VOSviewer* and *Gephi*. Firstly, relying on the WOS data described in the previous section, VOSviewer will create five term co-occurrence maps, extracting the main relevant keywords for each specific biennium. Secondly, Gephi will be used to analyse the network indicators, highlighting the centrality of specific words and their relationships.

The fields from which the terms will be extracted are the articles' titles and the articles' abstracts. The counting methodology applied is the *Full counting* one, meaning that all occurrences of a term in a document are counted (the Binary counting, instead, would have considered the presence or the absence only). For each network, the threshold of the minimum number of occurrences of a term has been set as 10, meaning that only the terms that appear at least 10 times in a document will be considered. For each term a relevance score is calculated and based on this score, the most relevant terms will be selected. Generally, after this step, the tool suggests considering a default 60% of all the most relevant terms. We will follow this suggestion. In the end, VOSviewer will generate a network composed by nodes (terms) and edges (co-occurrences in a same paper).

Since VOSviewer doesn't have the feature to calculate network indicators like nodes' degree or centrality, we would need another tool to calculate them, Gephi. We would need to export each network from the VOSviewer as a *.gml* file, in order to be able to open it with Gephi. This second tool will be useful also to represent graphically each network, leveraging on the network layout to show and highlight nodes' and edges' properties. Figure 10 below describes the whole process, from the data gathering to the analysis.

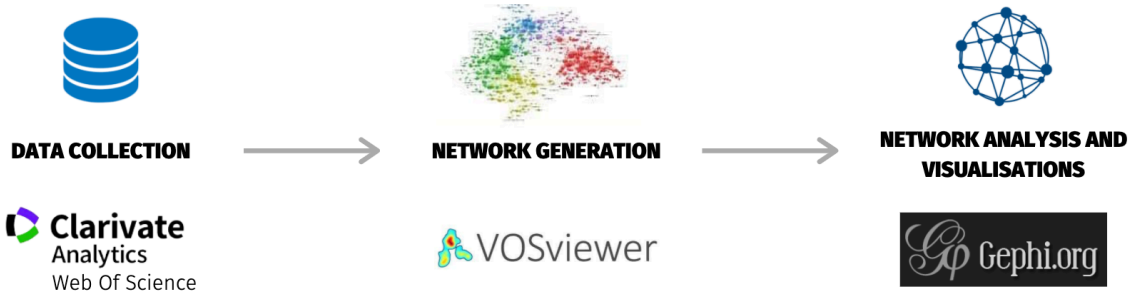

***Figure 11**. The project's process and tools.*

c. Co-word Network Analysis

Co-word Network analysis is a tool that is generally used to map a literature and research trends. This paper employs this method to investigate which topics related to AI, ethics and diversity have been relevant during the last 10 years across the academic literature. It is a technique generally used to (i) analyse the co-occurrences of key words, (ii) identify linkages, relationships and interactions between the topics researched and emerging research trends. The unit of the analysis is often the node, which indicates a term with a particular relevance within the network, where edges (or ties) cross and are connected to other nodes. The edges represent a connection between two nodes, identifying whether a term is likely to be present in a same paper or not. The stronger the tie between two nodes, the higher the number of papers which contain both the terms.

The main indicators that will be used to analyse the networks are: (i) the average nodes' degree (ii) the network diameter (iii) and (iv) the number of nodes and edges. Those will help to identify the networks sizes and properties.

4. Results and Visualisations

a. 2010 – 2011

The first co-word network (Figure 12) extracts 8 main keywords for the years 2010 and 2011, with a number of edges that equals to 24. 43 papers have been published during the specific time range. The average degree is 6, meaning that each node has, on average, six other nodes connected to. The main term in terms of degree is "analysis", identifying a starting concern related to AI and Ethics. To identify the network brokers, we leverage on the betweenness centrality indicator (used also to determine the nodes' size) and it can be seen that three terms are the broker ones: "algorithm", "analysis" and "approach". In terms of clustering the nodes, this network hasn't shown a specific need since in 2010 and 2011 so not much research has been yet developed.

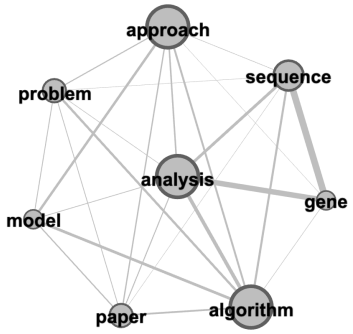

Figure 12. 2010 - 2011 Network

b. 2012 – 2013

During the next couple of years, an increment in terms of publications has been tracked: **54** (+25% growth). Simultaneously, also the number of nodes (11) and edges (49) became higher. The average degree now equals 8,9. As an interesting fact, we can testify how the term "problem" disappeared, shifting the focus from the issue identification to a more real concept, defined by the key term: ***data***.

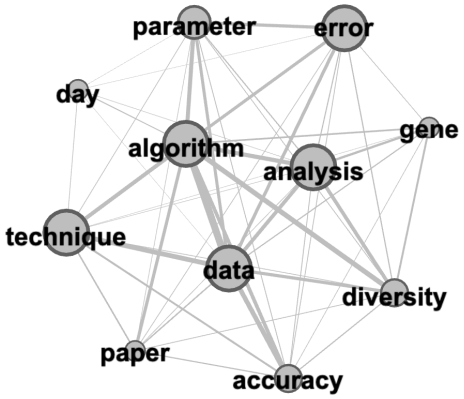

Figure 13. 2012 - 2013 Network

Also, the term *diversity* appeared in this second network as a relevant one, highlighting a new concern between 2012 and 2013 for ethical issues. Again, the nodes' size is determined by the betweenness centrality measure, identifying as network brokers the terms *algorithm*, *data*, *analysis* and *technique*. Clusters are not used here for the same reason described above.

c. 2014 – 2015

Between 2014 and 2015 the number of publications of interest has increased. There are 65 for this biennium. As described in figure 14, there are 87 edges and 15 nodes with an average degree that equals 11,6. The terms data and analysis have disappeared to leave space to *eda*, which means Exploratory Data Analysis. Two main clusters of nodes have been identified to describe the main area of each related term. The blue one can be considered as an **academic** area, highlighting the actual study of the AI and Ethics phenomena, also based on statistical terms like *population, parameter, simulation* or *model*.

The orange one, instead, seems to be related to issues (*problem*) concerning the proper *effectiveness* of algorithms/AI and *performances* on a **practical** and **professional** level.

The term diversity has been clustered automatically within the orange area. It should be seen as the main network broker (with the highest betweenness centrality - size) since it is the only one node that is connected with all the other components.

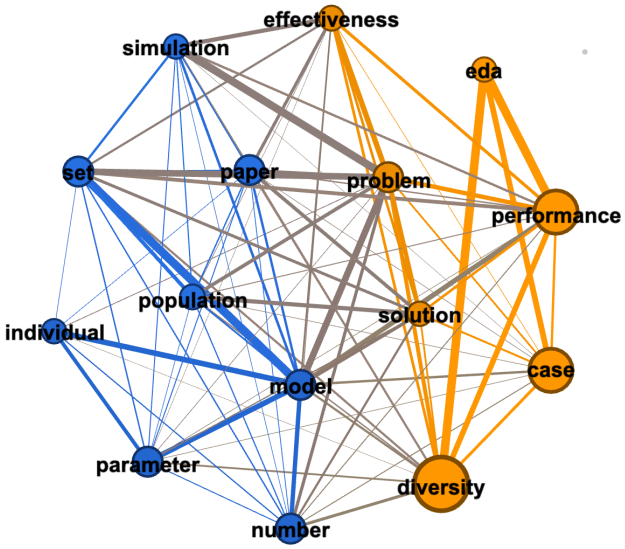

***Figure 14**. 2013 - 2015 Network*

d. 2016 – 2017

98 publications have been found for this biennium, +50,7 % than the previous set. There are 26 nodes and 231 edges, with an average degree that equals to 18,9. The two clusters discovered above remain. A new green one has been generated and includes all the terms related to the studied individual rights such as *diversity* - already present in 2014-2015 -, gender, and *privacy*.

This last one is particularly important since it can demonstrate concretely the birth of privacy studies, trying to balance the interests of personal data privacy and artificial intelligence and regulation (Humerick, 2017). From Figure 15 we can identify a growing concern for ethical implications related to AI and, specifically, to all the issues related to each individual. Not by chance the terms *user* and *person* appear in the network.

Within the blue cluster, the academic one, we can track the word *region* that could mean a geographical diversification of the study, since no common policies or regulations have been developed consistently for all the countries.

The network brokers can be easily identified as the terms *analysis* and *context*. This last one is important to be mentioned since it reflects the AI's property to be applied for various applications with diversified conditions.

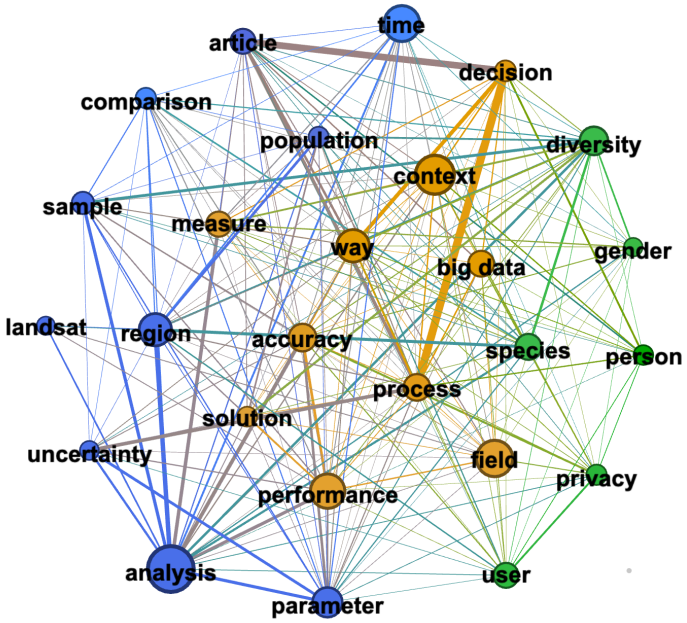

***Figure 15**. 2016 - 2017 Network*

e. 2018 – 2019

During the most recent two years analysed, 192 papers have been published with a growing rate of 98% in respect to the previous set. As expected, also the number of nodes and edges has increased, 63 and 1310 respectively, with an average degree of 41,2.

As Figure 16 describes, the clusters created have expanded, classifying new terms under the academic, professional and ethical collections. A new purple cluster has been generated and contains 5 terms that can be identified as future concerns related to AI and ethics issues like *education* and *accountability*. As mentioned in the literature review, in fact, during the last months a lot of frameworks and guidance documents have been developed by many organisations such as the European Union or the Chinese Government.

The term *discrimination*, falling under the green cluster, highlights how the issue generated by AI and algorithms biases could drive to unfair treatments, discriminating people by their gender and *ethnicity* (this term appears for the first time within this biennium).

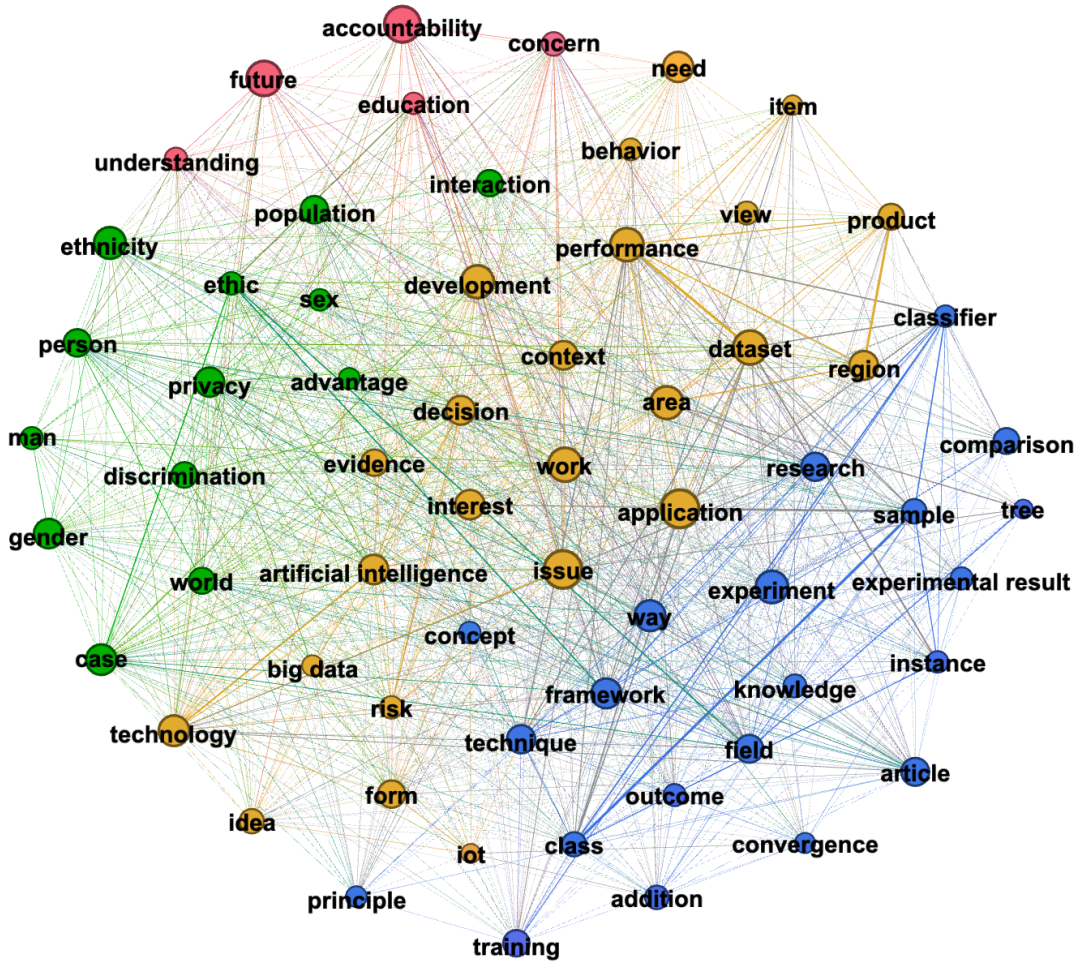

Figure 16. *2018 - 2019 Network*

f. Last 10 years aggregated

The last co-word semantic network analysed in this paper relies on all the articles and papers published between 2010 and 2019, aggregating all the papers downloaded from WoS. The number of all the publications equals obviously to the sum of all the biennium splits, a total of 452 documents.

As Figure 17 describes, the network generated has 164 nodes with an average degree of 76,5 and 6281 edges, having set a threshold of the minimum number of occurrences of a term such as 10 and, as for the networks above, considering a default 60% of all the most relevant terms.

Relying on the modularity of the network, calculated by Gephi, we have been able to create 5 different clusters, similarly to biennium networks. Each cluster has its own main terms related to the specific topic in terms of brokerage. The nodes size, again, is tied to the betweenness centrality of each node, in order to identify the most relevant ones.

The blue cluster contains terms related to the academic and research field. The orange one is tied to the professional and practical field. The red is a new one and collects all the terms related to actual tools, strategies and practices adopted when working with or studying AI. The green one is related to ethical topics. And the last purple one, again, as future concerns related to AI.

In the following chapter a deeper analysis and discussion will be presented, relying on network measurements and referring to existing literature to explain the co-word network properties and findings.

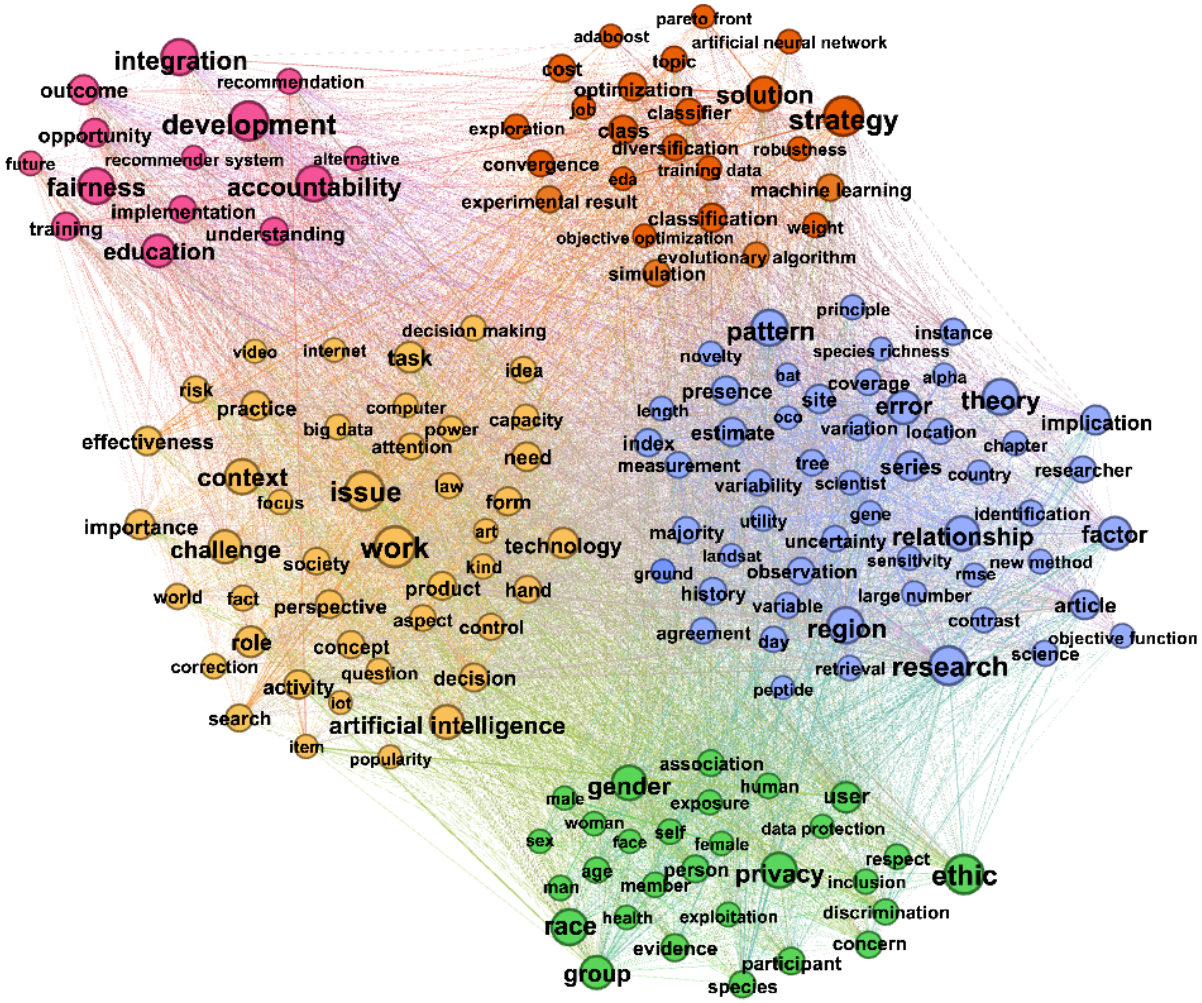

***Figure 17**. 2010-2019 Network*

5. Discussion & Analysis

a. Exponential increment of publications during the last 10 years

The image below synthetizes the concern growth related to AI and ethics between 2010 and 2019. In the previous chapter each network has been analysed individually, in order to get a few insights for each biennium. The discussion and analysis of those results will try firstly to find a narrative, identifying the main concerns for each specific time frame and try to understand how the focus has shifted from one topic to another during the years.

Secondly, the analysis will try to represent visually a graph over time where all the measures such as number of nodes, number of edges and average degrees have raised.

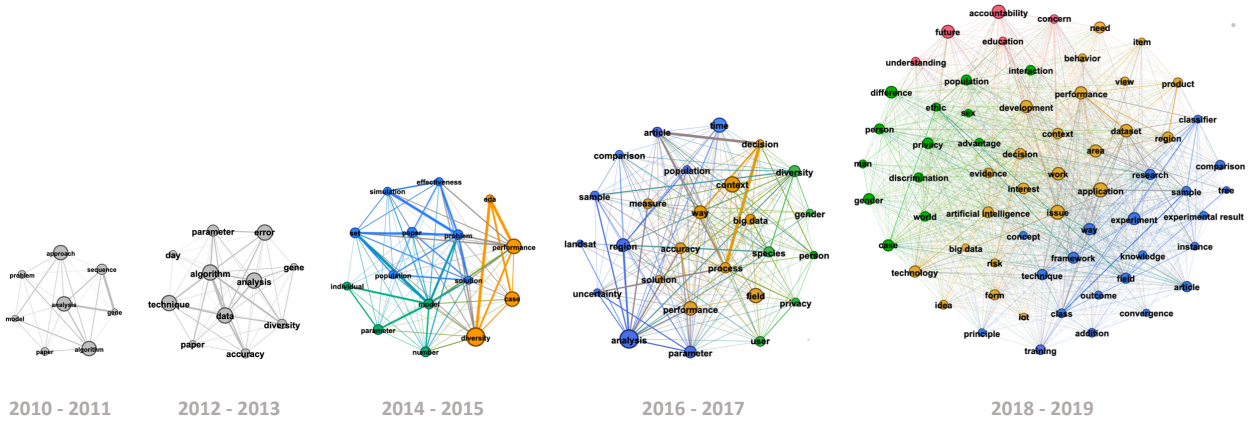

***Figure 18**. 5 biennium Networks, from 2010 to 2019*

Between 2010 and 2011, the terms *algorithms*, *analysis* and *approach* were the ones with the highest betweenness centrality, meaning being the brokers of that specific network. So, it is arguable to say that a key research question of the 43 papers analysed was trying to figure which approach should be used to develop good, ethical and fair algorithm. Different studies argue that operation and codes with accepted parameters does not guarantee ethically acceptable behaviour (Kraemer, et al., 2010). This is shown, for example, by profiling algorithms that inadvertently discriminate a specific set of people (based on their ethnicity) as happened in online advertisements according to perceived ethnicity (Sweeney, 2013). This was at the centre of the debate along the first years of 2010 decade. A typical conclusion of those discussions suggested a conceptual framework, based on critical analysis, aimed at informing future ethical inquiries, development, and governance of algorithms.

Nowadays this discussion has still a relevant importance since both academics and private companies are trying to shape the direction under a legal and regulatory perspective. But still, there is a lot to translate from principles into practices of digital ethics (Floridi, 2019).

During 2012 and 2013 the number of publications has increased by 25%. Still, the term *algorithm* gained a broker place, surrounded by other two important ones: *data* and *technique*. The first one finds its Latin roots in the word *Datum*, “thing”, neuter past participle of *dare* "to give", meaning a fact, given the basis for calculation in mathematical problems. Cambridge defines it as “*information, especially facts or*

numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer" (Cambridge Dictionary, 1946) and, as seen in the literature review, its amount has increased exponentially during the last decades.

The importance of this word is supported by its constant presence in the following co-word networks. Different studies, in fact, highlight the role of basic information to be crucial in building a strong, solid, ethical and unbiased AI. But how?

An answer can be found relying on the second term, *technique*, which lay the basics for a qualitative data collection. Since AI and algorithms are based on data and their results always depend on it, a quality data collection is the starting point to start building any system or software. Of course, even data can be biased, and a mismanaged use of them can bring possible paradoxes or unintended consequences. A clear example of this is related to facial recognition, like the Obama's example mentioned above, or the police use of live facial recognition technology (UK Governemt, 2019). Identifying people of colour faces relying on the analysis of mostly white faces database can bring to unethical consequences.

2014 and 2015 network started to shape a clear difference between the role of ethics in AI within the Academic sector against the professional one. This last one, contains the terms *performances* and *effectiveness*. While the academics seemed to be still studying and understanding the possible (and real) issues coming from AI, professionals and the private sector started trying to find possible ways to make profit out of it. In terms of innovation, large companies have developed their own venture groups investing in start-ups, running accelerators and building their own start-ups to ensure that they are leaders in AI driven solutions. Here below are a few companies that enhanced their customer service, increased productivity and improved products and services (Greenman, 2018).

***Figure 19**. Sectors and Firms investing in AI. Source: Greenman, 2018*

2016 and 2017 Network has evolved, generating a new cluster which includes all terms related to ethics and privacy. In the years before, the only word appeared in the network was *diversity*, now we can track others like *gender*, *species*, *privacy*, *person* and *user*. The last three ones refer distinctly to the Privacy world and the reason why they appeared is fairly related to the GDPR law promulgation.

On the 27th of April 2016, the General Data Protection Regulation (GDPR) published on the Official Journal of the European Union (European Commission, 2016), legislating on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. The law applies to any business that collects stores and uses data belonging to EU citizens. So, even those companies that were easily managing spreadsheets of client lists fell under the GDPR as collecting information assets.

Management of that information has become the more and more digital during the last decades and with the advent of the GDPR, many companies had to digitally transform themselves in order to be compliant with that law. Now, consequences of a badly managed data protection strategy are huge. Everyone will remember the largest known leak in Facebook history, when in 2018 Cambridge Analytica, after having harvested users data from the social network and sold millions of psychological profiles of American voters to political campaigns, had potentially violated the American election law, putting Facebook into its biggest crisis ever (The New York Times, 2018).

The GDPR has been vital move to protect Privacy as a human right. However, it will take time for data protection authorities to become effective, implementing and enforcing new resources, which are poorly allocated at the moment (The Human Rights Watch, 2018).

The last biennium network generated, compared with the first one, has undergone a huge increment in terms of number of nodes, connection and clusters. The 2019 has been tracked as the year with more terms related to future concerns of AI and ethics. Words like *accountability*, *transparency* and *understanding* became of central interest within the discussion. Recent studies tried to tackle social and ethical problems through the development of AI-based solutions, proposing essential guidelines to ensure that well-designed AI will more likely to serve the social good (Floridi, et al., 2020).

The following charts describe the main SNA KPIs to track the networks growth. Until 2016 and 2017, the growth is quite stable. Then between the last two biennia, a sudden raise has been tracked. Again, this is another explanatory way to say that during the most recent years, the publications and hence the study of ethics related to AI has increased exponentially.

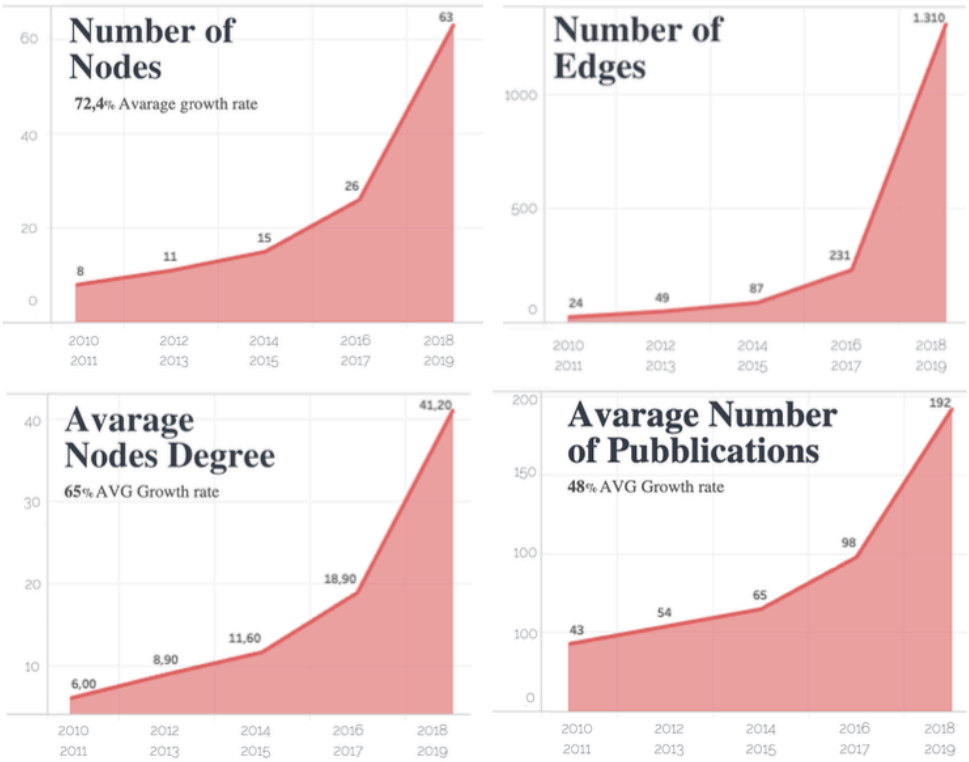

***Figure 20**. KPIs Growth charts*

b. The aggregated Network

As highlighted in the previous section, during the last ten years the studies related to ethics and AI have increased a lot. There have been numerous steps forward both in terms of publications and in terms of practical applications in our day to day life. Under an aggregative perspective, across the entire last decade, five clusters have been generated relying on the modularity of the network (Figure 17). The nodes represented are co-occurrences of key words extracted from all the papers without focusing on their year of publication, hence, the analysis will highlight their nature and the interactions between the topics researched and emerging research trends or topics. It is good to remember that the edges represent a connection between two nodes, identify whether a term is likely to be present in a same paper or not. As a first insight, the average degree is equal to 76,6. This means that each relevant term found in the network has, on average, 76 linked nodes related to different clusters. Figure 21 represents the average degree distribution which can be classified as a Normal distribution, having a high concentration of nodes within the centre and fewer within the extremities.

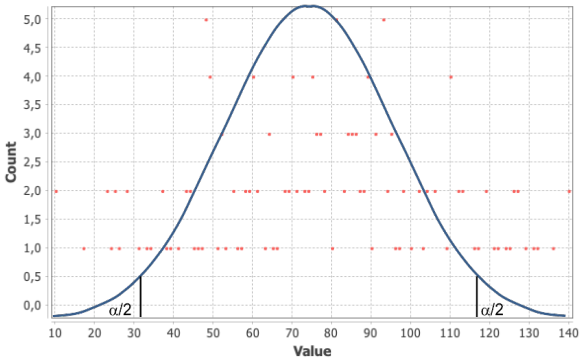

***Figure 21**. Average degree distribution*

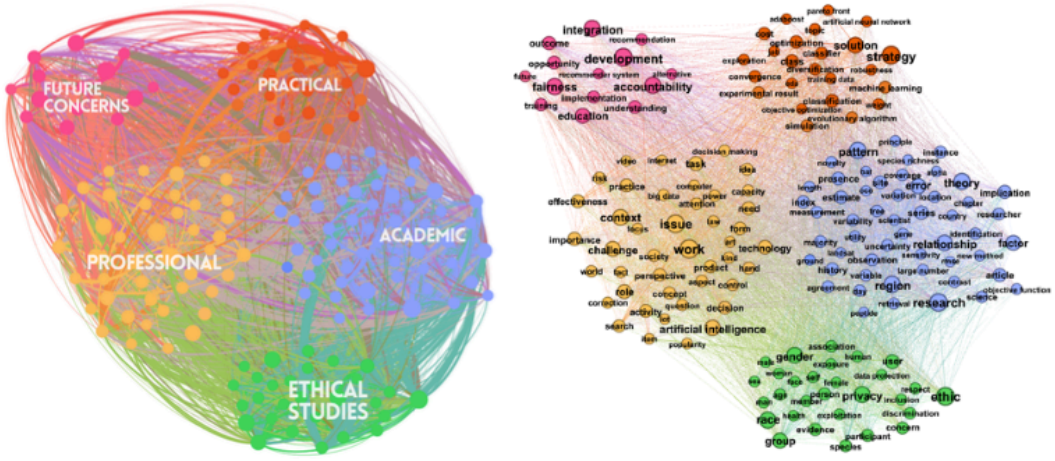

***Figure 22**. Aggregate network Clusters*

Each cluster collects a different number of nodes. The two greater ones are the Professional and Academic, followed by the ethical and then practical and future concerns. Each cluster size can fairly define the robustness of a specific field within the analysis. Professional and academic fields are in fact the bigger ones since, as we discovered before, their research has started years before the others. Future concerns, instead, is a brand-new and developing area, hence a slighter shape.

Keeping the focus on the edges, the following exercise will help us to define how each cluster is linked and shares co-words with another one. Analysing the key representative terms within each cluster, we will able to track all the connection of that specific node and evaluate its distribution. Figure 23 below, represents it visually: at the centre of each circle, there is the scrutinised node, while the lines, represent all its edges slit by cluster.

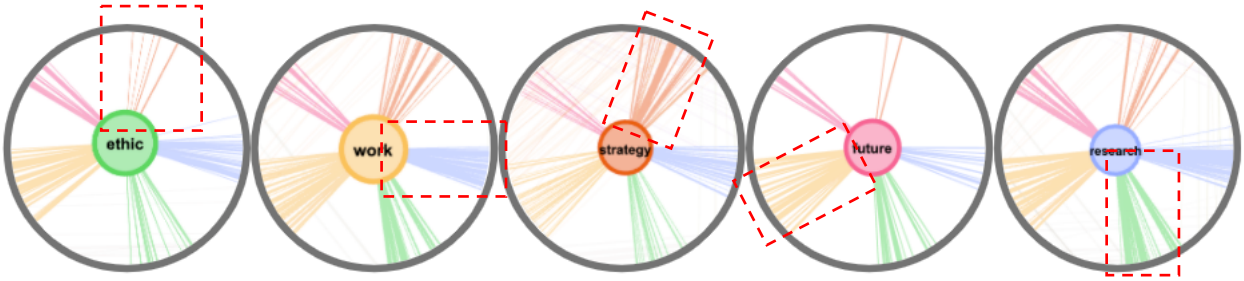

***Figure 23**. Main Terms linkages to other clusters*

Starting from the term *ethic*, which is the representative node for the green cluster, with a degree of 83 (higher than the average), the first insight that can be noted is that the connection with the strategic (and practical) cluster is very weak. This can be also noticed from an opposite perspective: the node *strategy* has a poor connection with the green cluster. The rationale behind this could be driven by the scarce number of publications studying these two topics together.

The *work* term, which is the representative node for the orange cluster, with a degree of 140 (the network's highest), has a robust connection with the academic field. This can be reasonably a very positive insight, since it arguably testifies a good collaboration and a shared soil between academia and professional papers, in order to tackle AI's ethical issues.

The term *strategy* which represents the red cluster, with a degree of 120 (higher than the average), present poor linkages in general with the other clusters. Instead, many connections between the nodes belong to its own field. The reason behind this could be determined by a lack of AI practical strategies to tackle ethical issues. Of course, being a new field of studies, this is understandable. Besides, it could be seen as a starting point of deeper studies which will come during the next few years.

The term *future* which represents the purple cluster, with a degree of 70 (6 connections less than the average), present a stronger affiliation to the work field rather than the academic one. This could mean that there are more papers discussing the future of AI and ethics leveraging on professional (instead of academic) terminology.

Finally, for what concerns the academic field, the term *research*, with a degree of 93 (higher than the average), presents a very strong relationship with the ethical studies. This is not the first time that philosophers try to figure which are the main ethical issues within the academia (Samuel, 1960), but certainly it is for Artificial intelligence.

6. Conclusion

Artificial Intelligence today cannot be truly considered morally intelligent. As seen in the literature review, it has evolved considerably in recent years, however it is not yet capable of making moral decisions. In 2018 an article was published in Nature, which carried out an experiment related to solving autonomous vehicle dilemmas and which did not lead to universal solutions (Awad, et al., 2018). Instead, it found that broad regions of the world displayed relative agreement, suggesting that our journey to consensual machine ethics is not doomed from the start, but is far from a real existence.

Having said that, over the past 10 years we have been viewing an exponential increase in articles related to the ethics of Artificial Intelligence. An increase that has mainly occurred during the last 2 years and which is likely to continue to grow (Statt, 2018). As we have seen in this paper, ethical problems that arise from an artificial "intelligence" are innumerable. From broken facial recognition used by law enforcement as a surveillance tool, to software that discriminate people of different genders based on respective incomes.

The conclusions that we can draw today from an analysis that identifies both key problems and the ways to solve them, are that there is a perceived convergence in the study of this subject. If academic and professional articles previously had two different paths which followed different rails, nowadays it seems that their directions are starting to approach each other, having the common good as their main goal, thereby minimising these ethical problems on a social level. Once the problems and the tools have been identified, a way must be found to solve these ethical problems, building and designing solutions that can prevent unwanted consequences. The utopia of a perfect society in a world without problems of any kind probably won't ever happen, but throughout time and improvements, we could reach it in an asymptotic way: the closer we can get to utopia, the more good things will happen. So, politics and governance, in this case, are no needed to solve a black or white issue. Instead, following a line that aims at reaching this utopia, they could have many ways to do it. And this is in a way more relaxing.

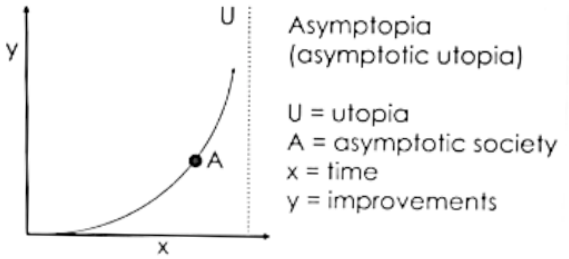

***Figure 24**. Floridi, L. (forthcoming) The Politics of Information*

Again, as Luciano Floridi remembers, in his papers, the near AI future should be shaped by a good **design** (Floridi, 2019), highlighting the importance of a good and fair development of algorithms (or AI systems). Defining innovation as a mix of three main ingredients: *discoveries*, *inventions* and *design*, our near future should rely on a portion of all of three. The discovery era has already happened, the invention one as well. Now we should behave as if we were in the design era, combining all the knowledge gained from the past and deep (and ethical) thinking on shaping a future that will be good for all. Leveraging not only on design, but on a constantly improved good design, made upgrade by upgrade.

7. Recommendations

During the first half of 2020, the world has been hit by the Covid-19 pandemic. One of the deadliest of the last few centuries. The state of emergency made us realize how enormous our frailties as people and citizens can be. However, our resilient nature has kept different sectors operational, thanks to the digital. From the smart working to the restaurants home delivery. This is to say how in difficult times, the problems that occur can be solved if we get together and find agreements. Moreover, we have seen how digital a very flexible tool can be, able to adapt to different types of situations.

Before the pandemic, our main concern as a specie was climate change. And it is still now. In this case it could help, for example, in the development of the philosophy of resources. Just notice how the circular economy and recovery have become gigantic sectors. We made great steps towards a global consciousness, but there is still a lot to do. Often can be noticed how even having the solutions in our hands, we struggle to reach our goals due to political problems. And this is a governance issue, also in terms of legislations as we have seen.

So, a first advice presented in this paper, whether you are a software developer, a politician or a doctor, is to rely on existing AI ethical frameworks to build a structured consciousness of the possible issues, thus being able order to avoid them and to protect everyone's human rights. Examples can be found within the literature review, such as the Beijing AI principles or the European Commission's Ethics guidelines for trustworthy AI.

The second advice is directly connected to the first one: leveraging on those frameworks, in fact, a set of new laws should be developed in order to regulate the market and its behaviour on a global level. This should be done consistently, going beyond geographical boundaries. Until a few years ago, each country had its own legislation. During the last decades, instead, we have seen how Europe, for example, has been capable to promulgate rules valid for many countries. Now, being the cyberspace without borders and rapidly growing, we need rules that should be fair and valid for any company whether its offices were in Dublin or in California.

Obviously, to arrive at a legislative system of this type, a lot of politics is needed. On the one hand, as mentioned above, in recent years it seems that populism has taken the reins, leading several countries to a stalemate. On the other hand, young people under age 30 constitute 2,2 per cent of the world's parliamentarians. The global proportion of MPs under age 30 has increased by only 0.3% points since 2016, from 1.9% to 2.2% (Inter-Parliamentary Union, 2018). So, in addition to the design of an ethical and legislative system relating to digital, it is also necessary to increase the interest of young people in the world of politics, in order to have leaders in the coming years aware of the problems analysed in this paper, and who have the intention to make the world a better place than it was found.

References

Ackoff, R. L., 1989. From data to wisdom. *Journal of Applied System Analysis,* Issue 15, pp. 3-9.

AI Now Institute, 2019. *Discriminating Systems: Gender, Race, and Power in AI*.. [Online]
Available at: https://ainowinstitute.org/discriminatingsystems.pdf
[Accessed 20 07 2020].

Awad, E., Dsouza, S. & Kim, R., 2018. The Moral Machine experiment. *Nature,* Volume 563, p. 59–64.

Bacon, F., 1597. *Meditationes Sacrae*. s.l.:Excusum impensis Humfredi Hooper.

Boisot, M. & Canals, A., 2004. Data, information and knowledge: have we got it right?. *Journal of Evolutionary Economics,* Volume 14, pp. 43-67.

Cambridge Dictionary, 1946. *DATA | Meaning in Cambridge Dictionary*. [Online]
Available at: https://dictionary.cambridge.org/dictionary/english/data
[Accessed 20 07 2020].

Corea, F., 2018. *AI Knowledge Map: How to Classify AI Technologies*. [Online]
Available at: https://medium.com/@Francesco_AI/ai-knowledge-map-how-to-classify-ai-technologies-6c073b969020.
[Accessed 2020 06 2020].

European Commission, 2016. *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*. [Online]
Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
[Accessed 20 07 2020].

European Commission, 2019. *Ethics guidelines for trustworthy AI*. [Online]
Available at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[Accessed 20 06 2020].

Floridi, L., 2019. *A.I. in society – Opportunities and risks*. IET London: Savoy Place, 2 Savoy Pl, London WC2R 0BL, s.n.

Floridi, L., 2019. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology,* Volume 32, pp. 185-193.

Floridi, L., 2019. What the Near Future of Artificial Intelligence Could Be. Issue 32, pp. 1-15.

Floridi, L., 2020. *The Green and the Blue: Naïve Ideas to Improve Politics in a Mature Information Society*. s.l.:Raffaello Cortina Editore.

Floridi, L. et al., 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines volume,* Volume 28, p. 689–707.

Floridi, L., Cowls, J., King, T. C. & Taddeo, M., 2020. How to Design AI for Social Good: Seven Essential Factors. *Science and Engineering Ethics volume ,* Volume 26, pp. 1771-1796.

Greenman, S., 2018. *Who Is Going To Make Money In AI?*. [Online]
Available at: https://towardsdatascience.com/who-is-going-to-make-money-in-ai-part-i-77a2f30b8cef
[Accessed 20 07 2020].

GSMA Intelligence, 2020. *Data and Analysis for the mobile industry*. [Online]
Available at: https://www.gsmaintelligence.com/data/
[Accessed 20 06 2020].

Holy Bible, n.d. *Proverbs 24:5*. King James Version ed. s.l.:s.n.

Hosanagar, K., 2019. *Talks at Google | AI Governance and Risk Management*. Palo Alto, s.n.

Humerick, M., 2017. Taking AI Personally: How the E.U. Must Learn to Balance the Interests of Personal Data Privacy & Artificial Intelligence. *Hein Online,* p. 393.

Inter-Parliamentary Union, 2018. *Youth participation in national parliaments: 2018,* Geneva: Courand et Associés.

Kraemer, F., Van Overveld, K. & Peterson, M., 2010. Is there an ethics of algorithms?. *Ethics and Information Technology ,* Volume 13, p. 251–260.

Kwiatkowski, S., 2018. *Entropy is a measure of uncertainty*. [Online]
Available at: https://towardsdatascience.com/entropy-is-a-measure-of-uncertainty-e2c000301c2c
[Accessed 20 06 2020].

McCandless, D., 2010. *TED Talks, The beauty of data visualization*. London, s.n.

Ministry of Science and Technology, 2019. *Beijing AI Principles*. [Online]
Available at: https://www.baai.ac.cn/news/beijing-ai-principles-en.html
[Accessed 20 06 2020].

Moore, G. E., 1965. Cramming more components onto integrated circuits. *Electronics,* 19 April.38(8).

PwC, 2017. *Artificial Intelligence in HR: a No-brainer*. [Online]
Available at: https://www.pwc.at/de/publikationen/verschiedenes/artificial-intelligence-in-hr-a-no-brainer.pdf
[Accessed 20 06 2020].

Sütterlin, C., 2020. *Upscaling Barack obama*. [Online]
Available at: https://twitter.com/Chicken3gg/status/1274314622447820801
[Accessed 20 06 2020].

Samuel, A. L., 1960. Some Moral and Technical Consequences of Automation—A Refutation. *Science,* 132(3429), pp. 741-742.

Statt, N., 2018. *The AI boom is happening all over the world, and it's accelerating quickly*. [Online]
Available at: https://www.theverge.com/2018/12/12/18136929/artificial-intelligence-ai-index-report-2018-machine-learning-global-progress-research
[Accessed 10 08 2020].

Sweeney, L., 2013. Discrimination in online ad delivery. 11(3), pp. 10-29.

The Economist, 2017. [Online]
Available at: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data
[Accessed 20 06 2020].

The Human Rights Watch, 2018. *The EU General Data Protection Regulation*. [Online]
Available at: https://www.hrw.org/news/2018/06/06/eu-general-data-protection-regulation
[Accessed 20 07 2020].

The New York Times, 2018. *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. [Online]
Available at: potentially violating American election law
[Accessed 20 07 2020].

UK Governemt, 2019. *Ethical issues arising from the police use of live facial recognition technology*. [Online]
Available at: https://www.gov.uk/government/publications/police-use-of-live-facial-recognition-technology-ethical-issues
[Accessed 25 07 2020].