Artificial Intelligence: theoretical, formative and communicative challenges of datification

La Inteligencia Artificial:
desafíos teóricos, formativos y comunicativos de la datificación

Inteligência artificial: desafios teóricos,
formativos e comunicativos da datificação

Dr. Víctor Lope Salvador
Lecturer of the Degree in Journalism
(University of Zaragoza)
https://orcid.org/0000-0002-9613-7671
Spain

Dr. Xhevrie Mamaqi
Associate Lecturer
Department of Economic Analysis
(University of Zaragoza)
https://orcid.org/0000-0002-4711-9792
Spain

Dr. Javier Vidal Bordes
Associate Lecturer
Department of Documentation Science and History of Science
(University of Zaragoza)
https://orcid.org/0000-0003-2400-4367
Spain

Abstract

This document explores, based on the recognition and definition of the new digital paradigm, the following topics: first, the need to catalog new skills and abilities for emerging professions in economics, business and communication; secondly, the recognition of a historical opportunity for the necessary theoretical and methodological innovation in Social Sciences and Humanities and, thirdly, the application of Artificial Intelligence to improve the quality of scientific publications. These three issues turn out to be nuclear in the opinion of the authors, insofar as the three affect the necessary renewal people training who are going to have to manage data of all kinds that affect the lifestyles of all. The work, after detecting the deficiencies in regulated training systems, raises the opportunities offered by the new digital paradigm in the fields of theoretical and scientific publications to face the inescapable challenges of new intellectual tools and new methods.

Key Words: *New digital paradigm; Artificial intelligence; DigComp; Communication; University education; Scientific reviews*

Resumen

El presente trabajo explora, a partir del reconocimiento y definición del nuevo paradigma digital, las siguientes cuestiones: la necesidad de catalogar las competencias y habilidades para profesiones emergentes en la economía, la empresa y la comunicación; en segundo lugar, el reconocimiento de una oportunidad histórica para la necesaria innovación teórica y metodológica en Ciencias Sociales y en Humanidades y, en tercer lugar, la aplicación de la Inteligencia Artificial (en adelante IA) para la mejora de la calidad en las publicaciones científicas. Estos tres asuntos resultan ser nucleares, a juicio de los autores, en la medida en que los tres inciden en la necesaria renovación en la formación de las personas que van a tener que gestionar datos de todo tipo que afectan a los modos de vida de todos los individuos. Por ello, este trabajo, tras detectar las carencias en los sistemas reglados de formación, plantea las oportunidades que el nuevo paradigma digital ofrece en lo teórico y en el terreno de la publicación científica para encarar los retos ineludibles de la nueva situación.

Palabras clave: *Nuevo paradigma digital; Inteligencia Artificial; DigComp; Comunicación; Educación universitaria; Revistas científicas.*

Resumo

O presente trabalho explora, a partir do reconhecimento e definição do novo paradigma, as seguintes questões: primeiro, a necessidade de catalogar novas habilidades e habilidades para as profissões emergentes na economia, nos negócios e na comunicação; em segundo lugar, o reconhecimento de uma oportunidade histórica para a necessária inovação teórica e metodológica em Ciências Sociais e Humanas e, em terceiro lugar, a aplicação da Inteligência Artificial para melhorar a qualidade das publicações científicas. Essas três questões acabam sendo nucleares na opinião dos autores, na medida em que os três afetam a necessária renovação na formação de pessoas que vão ter que gerenciar dados de todos os tipos que afetam os modos de vida de todos. indivíduos Portanto, este trabalho, após detectar as deficiências nos sistemas de formação regulamentados, levanta as oportunidades que o novo paradigma digital oferece no campo teórico e no da publicação científica para enfrentar os inescapáveis desafios de novas ferramentas intelectuais e novos métodos da nova situação.

Palavras chave: *Novo paradigma digital; Inteligência artificial; DigComp; Comunicação; Ensino universitário; Revistas científicas*

1. Introduction

The rapid development of datification, along with Artificial Intelligence (AI) in our contemporary lifestyle, has given rise to the construction of a new so-called digital reality, For now, the main difficulty, as far as research is concerned, lies in the imbalance between the fast-growing practical development of technologies and the shortcomings in the formulation of a theoretical base that can address the new analytical goals spawned by the use of such technologies. Researchers recognize that the procedures commonly used to collect information are insufficiently nuanced: current theoretical thinking fails to fully reflect real change and has been unable to undertake any in-depth studies into how people use, feel and think about internet and digitalization in general terms.

It is not easy to carry out empirical studies within a theoretical framework that is broad and solid enough to offer the necessary principles and premises in sufficient numbers. As Internet becomes an essential part of everyday life, it is clear that everything has changed on both a personal and collective level. Jobs need a progressively more versatile and multi-skilled workforce. Previously, work was far more strictly defined in terms of roles, responsibilities, tasks, schedules, etc. but now we have to be ready to work in a digitalized world where information and communication processes are performed at unprecedented speed. Technological development has the potential to create new jobs, increase productivity and boost innovation, investment and economic prosperity. However, in turn, there is also a notable concern for the growing control exercised by governments and large companies in all areas. Citizens and workers need to arm themselves with the digital skills and competences required for social inclusion and work placement in emerging sectors, and also to enable them to exercise their democratic rights and freedoms. These include the entire range of skills relating to information and communication at all levels.

2. Methodological principles within the framework of scientific, technological and social studies

The scientific, technological and social studies implemented in the early 1970s, particularly in the Anglo-American World, and also in France, provide a conceptual framework for tackling the rich network of connections and interdependences among the various objects under study that are of interest to different scientific disciplines - in this case, datification and AI technologies, together with the fields of knowledge involved, such as economics, communication, sociology and epistemology. However, it is imperative to work through some premises like the meaning of the term 'network', for example, as proposed by Bruno Latour, one of the best known theoreticians blazing a trail in this type of approach, suggesting that 'network' was conceived in terms of hybridization among humans, objects, technologies, discourse and nature. Such a proposal had the virtue of responding to the empirical finding that there is a flow of influences and causalities among things and subjects (Latour, 1993, p. 15) which gives shape to the reality of

social life, apart from the compartmentalization and separation of knowledge, things and discourses that scientific disciplines have defended since the days of Enlightenment. Nevertheless, recent work in social studies on technology such as those conducted by Australian Judy Wajcman (2017) show that Latour's primitive assumption of considering objects and humans to be on an equal footing has not been verified in reality. Wajcman shows that objects, including software, form part of the very human strategies that have to do with the use of time and that human desires are the first assumption that should be examined to attempt to understand the contradictions 9f contemporary technification.

In any case, when researching Information and Communication Technologies (ICTs), it is logical first of all to consider the warnings concerning the multiplicity of interconnections and implications that are issued with regard to the growing use of AI. As researchers Río and Velázquez (2005) argue, when talking about methodological procedures, we have to bear in mind the various approaches to scientific knowledge and the instruments used for observing reality (p. 43).

The multiple aspects and levels at which an analysis might be carried out makes it necessary to first determine which methodology to use. This means drawing up a plan in which the work is split into phases. The initial phase involves raising a problem that is considered to be interesting enough to initiate a process that might prove to be rather long and full of multiple derivations. How the problem is addressed will, in turn, depend on defining objectives, preparing questions and justifying the research (Río & Velázquez, 2005, p. 44). These three tasks are mutually influential and tackling any one of them will necessarily condition the other two.

What this article deals with is precisely the problem of certain challenges, both of an implicit and explicit nature for the subjects and for society with respect to datification and AI. Not surprisingly, sociological research has for some time assumed that social reality has two facets for the human race: subjective and objective and both elements need to be present in our task to provide mutual checks and balances (Giner, 1974, pp. 26-27).

From the multiple implications associated with these technologies the following three have been selected for this study as basic objectives: firstly, to update the repertoire of digital training skills in the professional analysis of big data; secondly, to explore the epistemological opportunities offered by AI in Social Sciences and Humanities; and thirdly, to study the possibilities this offers in enhancing the quality of scientific publications. These three objectives raise questions such as: What should be done with the data? Who should be responsible for analysing such data and what training should they receive? What guidelines should be established for future procedures as a result? What is the best set of digital skills and abilities for assessing the impact of the data that is stored, transmitted or studied on individual and collective lifestyles on various different levels? What are the implications for the theories and methods pertaining to Social Sciences and Humanities of big data-processing enabling correlations being established between completely diverse phenomena? How can AI help implement strict quality controls for the content of scientific publications?

More specifically, the main aim is to seek to identify those dimensions that need to be measured for competences associated with the tasks that are most in demand as a result of the technological revolution. There is also an attempt to develop the theoretical scope and perception of AI for Social Sciences and Humanities, and lastly, to facilitate quality control management with respect to scientific publications.

Both the objectives and the questions asked fully justify the approach adopted in this study in the current state of rapid technological development, with the confidence that it will spawn further research with a greater level of detail. As far as methodology is concerned, it is not a question of making predictions but rather of verifying the current gaps in theory and methodology by means of a combination of desk work (review of theoretical. empirical literature) and the use of statistical sources (collection and description of recent data), This gives rise to a perfectly feasible range of possibilities that can be logically deduced from current knowledge. If software engineers routinely carry out their work thinking about future solutions for the problems they envisage, Social Sciences would do well to adopt the same attitude by coming up with both problems and foreseea-

ble solutions to facilitate a fruitful dialog with other scientific and technological areas. Obviously, this attitude differs from the widespread belief in technological determinism that is often adopted uncritically.

3. Working procedure

3.1. Digital paradigm

The absence of applied research along with the diversity of definitions hampers the work of identifying common criteria for assessing digital skills and abilities. Finding a single definition with respect to the digital paradigm may well prove to be an arduous task. The impact is so broad and profound that the difficulty of assembling the whole puzzle of this new paradigm is quite impressive as each area and scientific discipline has adapted in accordance with its own rules and traditions.

A scientific revolution occurs, according to Kuhn (2008), when scientists find anomalies that cannot be explained by the universally accepted paradigm and the discipline becomes embroiled in a state of crisis. In the terminology used by Kuhn, a scientific revolution or a paradigm shift occurs precisely as a result of adapting to new realities that we wish to learn about.

A general but precise definition of *digital paradigm* can be found in IGI Global's *Dictionary Search*:

Pattern of reality that results from a sampling process encoded in binary language. Unlike the analog paradigm, characteristic of a physical reality composed of complex signs that are not yet computable, the digital paradigm is reflected in a virtual reality composed entirely by computable finite groups of just two different signals (0 and 1).

It is quite difficult to separate the theoretical concept and the technological tools that gave rise to them on the basis of the definitions put forward in the literature. Heeks (2016) conceives of the problem just at the moment when this paradigm is starting to be widely adopted in the following terms:

...a digital development paradigm which conceptualizes ICT not as one tool among many that enables particular aspects of development, but as the platform that increasingly mediates development.

The same author details three phases of development for the new digital paradigm from a chronological perspective: 1) the pre-digital paradigm, 2) the ICT4D paradigm, and 3) the development of the digital paradigm (Figure 1).
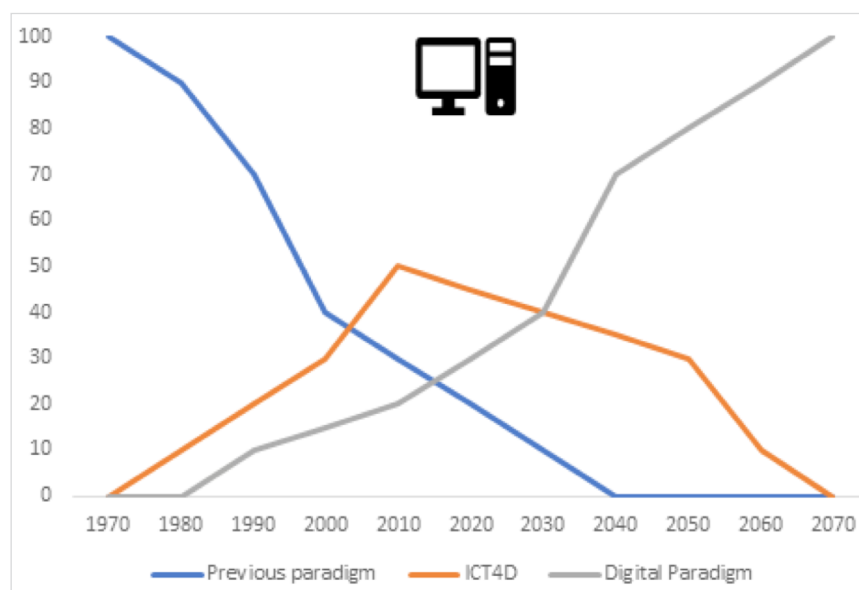

Figure 1: Development and chronological shift of digital paradigm.
Source: Adapted from Heeks (2016) original.

The pre-digital paradigm has lasted 50 years and has conceptualized the separation between new ICT and development (Heeks, 2009). The ICT4D paradigm began in the 1990s and promoted new conceptualized ICT as a useful tool for development. The paradigm emerged as a symbiosis between the onset of the generally availability of the Internet as a tool that is the basic condition for generating benefits and uses and the objectives for millennial development. And lastly, the digital paradigm, which took off at the beginning of this century, will continue over the years ahead and reach its plenitude at the end of the 21st century.

It is impossible to understand the new technological revolution or the consolidation of the digital paradigm without reviewing the evolution of the cellphone from its beginnings until the advent of 5G. First, we need to run through the history of its predecessors (1G, 2G, 3G, and 4G) to understand how our current reality was built up from these earlier generations.

Although new ICT go all the way back to the 1940s, it is really in the 1980s when 1G started to be used for mass wireless communication. This enabled us to make simple calls from one cellphone to another. The data transfer rate was then about 0.01 MB per second.

Ten years later (precisely in 1991), 2G technology was introduced, offering greater security through the use of digital encryption, unlike analog signals, and a faster rate of transfer of up to 3.1 MB per second, which was triple the speed offered by 1G technology. It also enabled regular text-only SMS messages to be sent.

Seven years were to pass before 3G technology was achieved and it was precisely in 1998 when the smartphone revolution kicked off, offering speeds of up to 14.4 MB per second and connecting mobile devices to Internet.

Almost a decade was to go by before the current standard, 4G / LTE, came into being in 2008. The most notable change was the incredible leap in terms of speed to 300 MB per second, which enabled the users to take part in activities like the transmission of high-definition audiovisual content.

In the last decade all effort regarding ICTs has focused on launching 5G. This offers the same basic features as its predecessors (text messaging, cellular voice calls and Internet connectivity), largely bolted on to the core 4G LTE Long Term Evolution) technology. Nevertheless, 5G technology stands out for four notable changes: bandwidth, which is expected to reach 1 GB per second; latency times - delay in the real process of data transfer – have been reduced to less than a millisecond; greater energy efficiency and an exponential increase for network connectivity.

The relevance of speed is self-explanatory, without the bandwidth, or size of the data packets transferred, being significant at any given moment. Increased energy efficiency is relevant since it significantly reduces costs and increases battery life, while also reducing CO_2 emissions (Lope, Vidal & Mamaqi, 2018).

That said, the most important change is that of increasing network capacity since this forms the bedrock for the development of the Internet of Things (IoT). This all involves the creation of new infrastructures and a spectacular amount of growth in potential demand on the part of all types of users.

Mobile technology					
Characteristics	1G	2G	3G	4G	5G
Deployment	1980	1991	2004-2005	2006-2010	Desde 2020
Band width per second	0,01MB	3,1MB	14,4MB	300MB	1GB
Tecnology	Analogical	Digital	IP Techno-logy	IP analogi-cal and uni-fied LAN/ WAN/WLAN/ PAN	4G+www
Service	Single voice No security	Digita voicel, SMS, autentica-tion ect	Integrated high quality audio, video and data	Dynamic access to informa-tion, device variables	Dynamic access to information, device varia-bles with all capabilities

Table 1: *1G vs 2G vs 3G vs 4G vs 5G: basic features.*
Source: Authors' original work.

As can be easily discerned, the development of the new digital paradigm is marked by the 5G revolution and mobile technology.
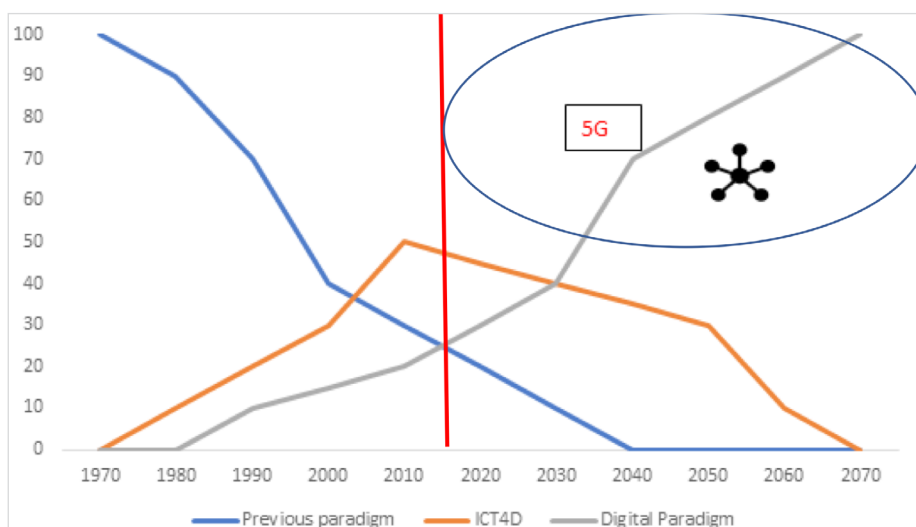
Figure 2: *Digital Paradigm and 5G.*
Source: Adapted from Heeks' original work (2016).

Apart from being faster and more stable, and reducing latency from 1 to 3 milliseconds, 5G technology will enhance user experience (UX) by allowing mass use of virtual reality, augmented reality, IoT, tactile Internet (with applications in medicine, for example), automatic vehicles, and smart cities, among many other options. Of course, the question of whether we are prepared to face these drastic changes inevitably arises. Digital reality involves a permanent, ever-increasing link between users, companies, governments, etc. There should be full integration on the part of new digital professionals to this new reality.

3.2. Emerging digital professionals: skills and abilities

In the digital business world, one of the greatest challenges raised by this new situation is that of making the most of the huge quantity of data - *big data* – to establish patterns for immediate reaction, the capacity to make predictions and enhance user experience. This all has a direct impact on the matter of digital competence.

In this respect, in 2006, the European Parliament and the EU Council (*European Parliament and the Council, 2006*) defined digital competence as follows:

Digital competence involves the confident and critical use of Information Society Technology (IST) for work, leisure and communication. It is underpinned by basic skills in ICT: the use of computers to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in collaborative networks via the Internet.

In this definition the use of ICT has a prominent role in four large areas: for information, work, leisure and communication, subsequently highlighting the following skills: using computers, retrieving, reproducing, and saving information, skills for exchanging information and communication via Internet. The European Digital Competence Framework for Citizens sets out five macro-areas: i) information and data literacy, ii) communication and collaboration, iii) digital content creation, iv) safety, and v) problem-solving. Included in basic digital skills are technological skills, creativity, critical thinking and evaluation and effective collaboration (Ala-Mutka, 2011).

In practice, it is difficult for emerging digital professionals to find solvent scientific and institutional references for rigorous intellectual thinking and stringent ethical demands, a catalog of internationally recognized specific skills and abilities (Russom, 2011).

Among today's various digital professionals – taking into account the most immediate needs for companies and their growth expectations – the five most requested profiles in recent years are Digital Marketing Manager, Digital Community Manager, Communication Manager, and Digital Analyst, as well as other emerging professional profiles that are configured on the fly in accordance with specific needs (Zbigniev, Chand Seal, Leon & Wiedenman, 2017).

The emerging discipline of Data Science covers the analysis, display and management of large sets of records (where 'large' in this context means many millions or billions of entries) The digitalization of all kinds of information, increase in the

number of sensors and cheap storage have combined to generate huge amounts of data that are of interest to the Social Sciences sector and also to businesses (De Mauro, Greco, Grimaldi & Ritala, P., 2018).

Since the past decade, several authors have described the statistical work of data scientists as the sexiest job of the next 10 years (Granville, 2014). Forbes magazine describes the role of the data scientist as being the new concert in technology; and that Data Science is where geeks go (Marr, 2016). Many managers of large companies have already warned about the possible shifts that were going to take place but it is only recently in the middle of this decade when his premonition has proved to be correct (Miller, 2014). According to the EPYCE report database (2017) significant changes for professional technology families have come about since the biennium 2016-2017 and with forecasts for the next four years, until 2021. As can be seen in Table 2, many of the professionals that were linked to Technology in 2016 have disappeared or their name has been adapted to evolving reality, with the profession of *Data Science Analyst* emerging in 2017.
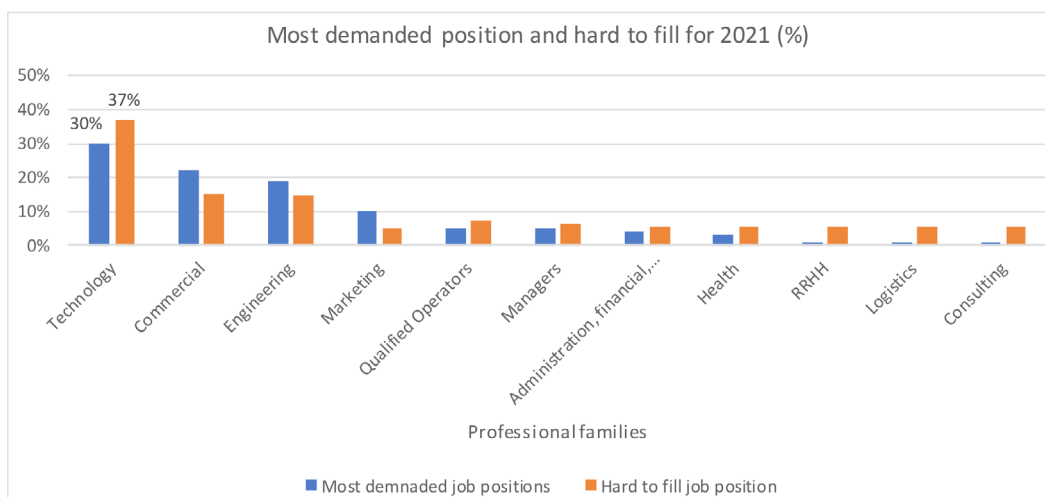
Profesional families: Tecnology. Year 2016
Big Data (It keeps)
Web analyst developert (Renamed as Web Developer)
Ecommerce (Renamed as Ecommerce Developer Specialist)
Application Specialist (Disappears)
Integration Specialist (It keeps)
Information System Specialist (It keeps)
R&D Specialist (It keeps)
Project Leader (Disappears)
Computer Programmer (It keeps)
Web Programmer, iOS, Android (It keeps)
Project Manager (It keeps)
Responsible for Cibersecurity (It keeps)
Communications Technician (Disappears)

Professional family: Technology. Year 2017
Big data
Data Science (Emerges)
Web Developer
Multimedia Applications Developer
Ecommerce Development Specialists
Integration Specialist
R&D Information System Specialist
Computer Programmer
Web Programmer, Ios, Android
Project Manager
Cloud Manager

Table 2: *Technology and adaptation of new professional positions.*
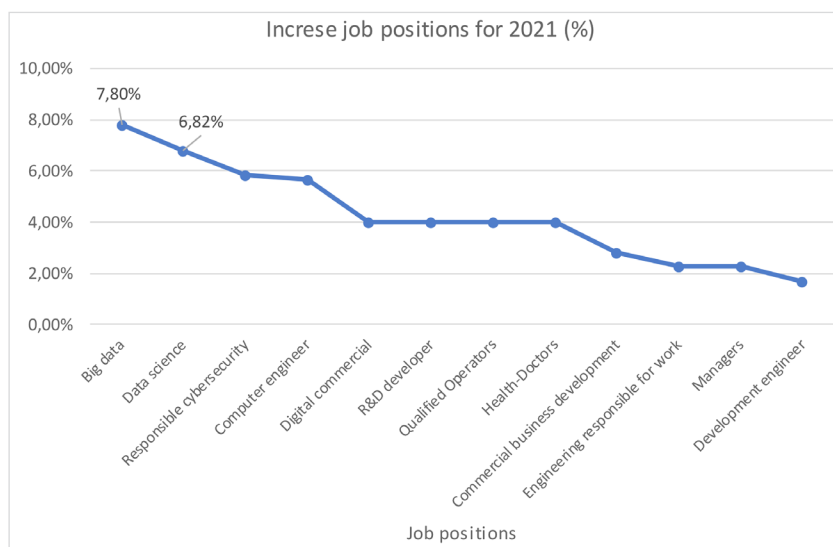*Source: EPYCE report (2017) and own elaboration.*

In the immediate future, the professions in greatest demand and most difficult to cover, as can be seen in Graph 1, concern jobs involving technology and data processing.

Gaph 1. *Professional families with Jobs in greatest demand and hardest to cover.*
Source: Own elaboration (EPYCE database, 2017).

The positions of *Big Data* and *Data Science* are the most requested, according to forecasts for 2021, ranking first and second in a total of 32 positions of different professional families (in Graph 2 the top 12 positions are shown for forecasts for 2021). There are a total of 12 positions that make up almost 42% of the increase in demand in the labor market for 2021 whereas 20 positions only increase by 0.05% and some completely stagnate like the positions pertaining to Human Resources and Technical Marketing Engineer.

Graph 2: *Forecast for increase in Jobs in greatest demand*
Source: Own elaboration (EPYCE report database, 2017).

As can be observed, the most recent profiles have emerged from the discipline of Data Science. The *Data Scientist* is, in turn, the profession that best aligns with the digital paradigm.

In a search for jobs as a Big Data Analyst, only in one of the nation-wide technological employment portals is it revealed that in the past month the need for a total of 621 jobs related to these professions is announced (Figure 3).
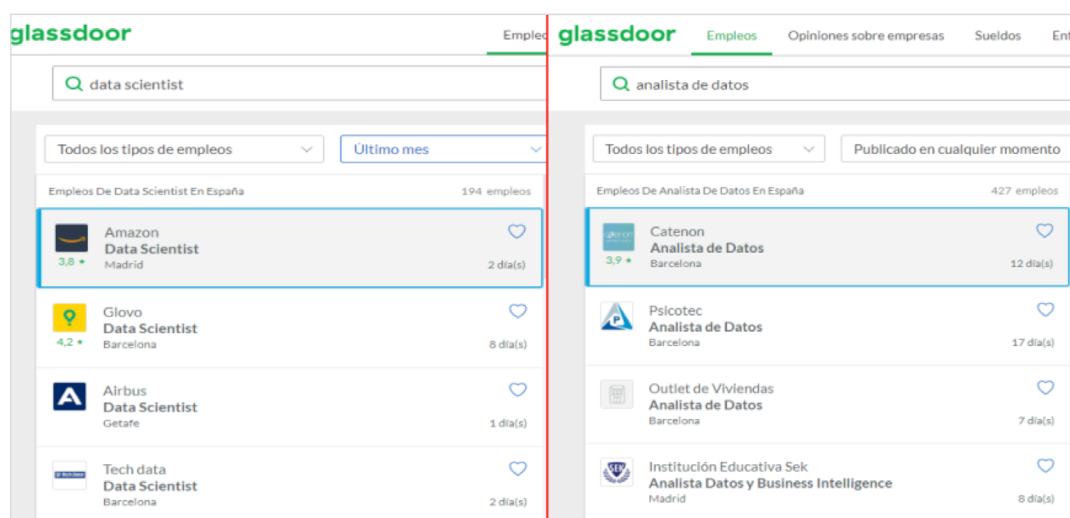
Figure 3: Jobs for Big Data Analysts and Data Scientists announced in the past month.
Source: Authors' original work (October 2019).

However, once the need has been detected, its development and cataloging require rapid progress (Mamaqi, Miguel & Olave, 2011). Data scientists use specialized techniques to examine these troves of information to discover new ideas and create new value (White, 2012). In this profile, skills from different spheres and specialties of knowledge are grouped together, such as mathematics, statistics, economics, computer programming, business administration and advertising (Naur, 1966).

For the moment, these professional profiles do not require specific training at a Spanish university (most studies are for computer programmers or part-time or non-integral training), although unregulated courses can also be found. In practice, these professionals are trained on the fly and from the accumulation of shared experience. The best contributions in this regard are to be found intermingled with loose contributions from independent professionals, consultancies, etc., without a line of qualifications being set up to enable a complete teaching-learning process to be defined from start to finish and to ensure that it can be adapted to the digital jobs required in the market (Mamaqi, Marta-Lazo & Pérez, 2019). Data Science has, for many, proved to be a discipline of recent creation, with this role being assigned in the digital age in 2008 to Patil and Hammerbacher, data

analysts from Facebook and LinkedIn respectively. Both predicted the importance of these professional profiles, immersed in huge amounts of data, organizing them, drawing conclusions so as to improve company operations and give them a key role in the decision-making process in various areas (Russom, 2011). However, its origin is much more ancient since Tukey (1962 and 1977, Naur, 1966) should be remembered when explaining the evolution of mathematical statistics. Thus, Data Analysis needs to be defined for the first time as a procedure for interpreting them. The aforementioned authors proposed ways to plan the collection of data to make analyzing them much easier, more precise or accurate and ways of suggesting hypotheses to be tested in statistical models. In 1996, the term Data Science was used for the first time at a conference entitled "Data Science, Classification and Related Methods", organized by members of the *International Federation of Classification Societies*, where the statistical work was described as a trilogy consisting of re-collection of data, data analysis and modeling, and decision-making, requesting that statistics be re-named Data Science and statisticians Data Scientists. And at the beginning of the century, Data Science was introduced as an independent discipline, expanding the field of statistics to include the progress made in data computing. In 2012, at another international science event (The International Council for Science: Committee on Data for Science and Technology) CODATA (The Committee on Data for Science and Technology) was set up, redefining Data Science as Data Science and Technology.

De Mauro, Greco, Grimaldi & Ritala (2018) group the positions around Data Science professionals as: i) expert in data tools, ii) programmer, iii) quantitative statistician/analyst, iv) researcher v) data hacker vi) auditor vii) data protection and ethics manager and viii) data and strategy manager.

From what has been described above, at a minimum, it can be seen that these professionals more than comply with the basic digital skills and abilities in regulated education, with a certain level of perfectionism being achieved in Master's and PhD degrees. Conceptually, the skills and abilities required of Data Science professionals can be divided into two groups:

1. Technical and methodological skills and abilities related to the use of technological, statistical and analytical tools that are suitable for studying huge databases. This group requires multidisciplinary training that can mold the different skills and abilities associated with specific knowledge pertaining to data analysis.

2. The capacity to enable value transformation using management techniques and knowledge of a specific business domain. This group includes soft skills, mainly in the area of communication.

With reference to the first group we have Cleveland (2001), who established six technical areas making up the field of Data Science: multidisciplinary research, models and methods for data, data computing, pedagogy, evaluation of tools and theory. Adapting these six areas to the skills of the first group within the professional Data Science family requires three positions: a Data Scientist, a specialist in Big Data and one in Data Analysis.

Based on these technical and practical contributions, in the following table we can see the list of requirements for top-level skills and specialties.

Professional family:	Professional positions		
Technology	Data scientist	Big Data Analyst	Data Analyst
Education Level	Master and/or PHD	Master and/or PHD	Master
Specifics competence	Advanced level		
-Knowledge of statistical programs	R and SAS	Statistical programs	Programming R and Phyton
-Communication language	Phyton, Hadoop platform, SQL	R, SAS, STATA, SPSS, etc.	Data programm management
-Data manegement	Extract text data from social networks, video and audio.	Extract, transform and data analyze.	Handle different raw data formats. Perform necessary to transform between different data formats.
-Type of data	Unstructured data	Unstructured and structured.	Unstructured and structured.

Professional family:	Professional positions		
Technology	Data scientist	Big Data Analyst	Data Analyst
-Análysis of data	Extensive and solid knowledge of mathematical methods to develop algorithms, statistics analysis and extract information properly. Understand fundamentals of data analysis: Data metrics. Develop digital indicators performance for different business models. Deep and solid knowledge of mathematical-statistical methods.	Extensive and solid knowledge of mathematical methods to develop algorithms,statistics análisis and extract information properly. Understand fundamentals concepts of data análisis and data metrics. Develop digital indicators performance for different business models. Deep and solid knowledge of mathematical-statistical methods. Extensive knowledge about different key performance indicator (KPi-s).	Deep and solid knowledge of mathematical methods to develop algorithms and statistics analysis to extract information properly. Understand fundamentals of data analysis: Data metrics. Elaborate numerical and text data. Develop indicators and synthesize data information. Extensive knowledge about different key performance indicator (KPi-s)
	Business and organizational knowledge. Data control at all stages of the company.	Business and organizational knowledge.	Business and organizational knowledge.
Capacities			
Problem resolution	Problem solving through innovation. Remove obstacles from your business scope scope.	Identify and solve specific problems by data anlysis.	Solve visualization and data processing problems.

Professional family: Technology	Professional positions		
	Data scientist	Big Data Analyst	Data Analyst
Critical thinking	Ability to relate data with the business model.	Data comprehension.

Select relevant data for the company. | Intuition on how to relate data with qualitative vs. quantitative appropriate analytical methods. |
| Creativity | Provide valuable and contrasted information based on data tracking, promoting strategic guidelines in the short and long term.

Test and validate ideas.

Mark strategy guidelines according to data. | Select appropriate, qualitative vs. quantitative techniques.

Collect and analyze the data properly. | Provide trends and establish guidelines through data analysis.

Explore the data to mark the difference with competence. |
| Decisions taking | Promote innovation in the context of organization.

Promote changes in the company / organization based on data contrast scientifically. | Align the analytical elaboration of the data to the objectives of the company. | Understand the sense of business.

Adaptate the specifics knowledge to business objectivs. |

Professional family:	Professional positions		
Technology	Data scientist	Big Data Analyst	Data Analyst
Digital comunications	Domination oral-written digital lenguage.	Domination oral-written digital lenguage.	Domination oral-written digital lenguage.
	Communication skills to prepare summaries and reports.	Communication skills to prepare summaries and reports.	Communication skills to prepare summaries and reports.
	Ability to provide data as understandable ideas for the company executive.	Ability to provide data as understandable ideas for the company executive.	
Conduct	Profesional ethics.	Professional ethics.	Professional ethics.
	Personal integrity.	Personal Inegrity.	Personal integrity.

Table 3. *Professional competences and capacities: Data scientist vs. Big data analyst vs. Data Analyst Source: Own elaboration (July-October 2019 period).*

3.3. Theoretical opportunity in the hands of mass data

As Bertalanffy, Ross Ashby, Weinberg et al. (1978) recalled in the 1970s, scientific and philosophical thinking has historically swung between systemic and holistic positions postulating, in the words of Aristotle, that the whole is greater than the parts (p.29) and that the Cartesian and positivist conceptions emerging in Europe in the 16th and 17th centuries gave rise to the industrial revolution. These latter conceptions are seen to become efficient once the problems have been broken down – fragmented, separated and distinguished – in such a way that complex phenomena can be analyzed by reducing them to their elementary parts and processes (p.31).

We are compelled to recognise the fact that, from now on, the processing of big data can serve both conceptions at the same time, and without implying any theoretical or methodological contradictions (Lope, 2018, p.146). Studies based on

big data can establish correlations between phenomena that, in principle, are not thought to be connected, thereby illuminating new systemic visions. In turn, analysing such data will enable objective causal connections to be detected. Hence, thanks to technological progress, the two most significant trends of Western science are seen to be facing a theoretical reformulation with far-reaching consequences.

It is therefore imperative to recognize the new epistemological situation – new episteme derived from the latest version of *téknē* – and take advantage of it in all its potential for scientific research, particularly in the field of Social and Legal Sciences and Humanities.

That said, certain metaphysical inductions in the use of big data from a system or holistic standpoint should be avoided. As Google's ex-data scientist Stephens-Davidowitz (2019) usefully points out, the temptation to make predictions after detecting simple correlations of events usually leads to failure as it should be borne in mind that many correlations are nothing more than fortuitous coincidences and do not prove any causality whatsoever. The solution is not always a case of finding more big data. A special ingredient is often needed for it to work best: human judgment and small-scale questionnaires, i.e. something that we could call small data (Stephens-Davidowitz, 2019, p. 253).

This theoretical innovation we are trying to outline cannot ignore the fact that the hugely expanding incidence of innovations in people's everyday life has ethical implications. We are apparently offered things and satisfactions when, in practice, such things and satisfactions require us to adopt a different lifestyle. The imposition is received through our affective channels, as Mark Hunyadi (2015) cleverly points out. In this way it appeals to us, or at least something about it appeals strongly enough for us to accept the rest of it that is less appealing. According to this model, a system gradually manages to eventually impose a lifestyle that no one has explicitly wished for, that perhaps nobody would have wanted if they could have chosen outright (p.70).

It is essential to place something in the theoretical middle that we tend not to want to see. This is because the technique is never neutral from an ethical stand-

point. It is not true that a tool's goodness or evil depends solely on the use – whether beneficial or destructive – that is made of it. (Mamaqi, Marta-Lazo & Pérez, 2019).

Judy Wajcman (2017), the specialist in sociological studies on the impacts of technology on people's lives does not mince her words, saying that the time has come to question the euphoria for speed and the technological impulse to achieve it, by harnessing our inventiveness to gain control of our time for a bit longer (p. 257).

Hence, the need to assume, both regarding theory and methodology, the initial need to carry out a thorough analysis, just like any scientific task, applied to all types of tools and services, irrespective of whether they are real or virtual. Algorithms, robots, and applications should be subjected by the scientific community to rigorous studies that evaluate their characteristics taking into account that we are dealing with complex texts we engage with and which participate in the configuration of our subjectivity in registering what is real, what is imaginary and what is semi-cognitive and which, occasionally can even have a symbolic dimension (Lope, 2018).

The theorization of communication processes itself requires the rethinking of such tenets since, by and large, the Shannon and Weaver Model of Communication now only applies to the problems for which it was originally conceived, i.e. for transferring information from one machine to another. On the other hand, communication processes among humans – whether employing secondary, tertiary, or quaternary means – are far more complex than the theoretical simplification proposed by Shannon and Weaver.

It is no longer acceptable to entertain any model of human communication that does not incorporate the recent discoveries made in neuroscience. Among such discoveries is the one detected by Freud and confirmed by current research that the origin of our decisions and functions – communication being one of them – are unconscious. From this we can derive a very important notion: the idea that a subject already exists prior to the communicative act is no longer defensible in any circumstances. If unconscious desires do exist, this idea that there is a subject that is fully aware when taking a decision to issue a message is fallacious. The Internet of Things does just that: it amasses mountains of data on the subjects' unconscious activities.

In fact, the most important thing about human communicative processes, namely meaning, is the most difficult to pin down in theoretical terms. Gonzalo Abril (2005) reminds us that meaning is immeasurable and paradoxical. Since, unlike what happens with the information processed by IT, it is not possible to generate a metalinguistic discourse for meaning, because no external discourse exists outside meaning itself (p.36). And for the real experience of each subject with signals, images, speech, and objects, the really decisive thing is meaning. As explains neuroscientist Antonio Damasio (2018), the merely cognitive level must be distinguished from the emotional level even though both things are practically inseparable in the individual's subjective experience. And meaning, according to Damasio, should be understood as the real value «felt» by the human being from life's consequences, good or bad (p.31). The alliance between Data Science and Neuroscience should provide ways to integrate meaning and subjectivity in science.

Back in the early 1970s the philosopher and sociologist, Julien Freund, explained the dangers involved in ideological deviations that were undermining the scientific nature and rigor of the Social Sciences and Humanities. In the first two decades of the 21st century, there is no sign of improvement but rather of uncritical accommodation within formulaic expressions – largely generated as a result of political correctness – which are offered as valid concepts without even devoting any intellectual effort to demonstrating their theoretical validity (1975). What Freund (1975) was saying now takes on undisputable value and relevance: The human sciences specialist does not only work with objective facts but also with opinions, beliefs, meanings. Their object consists of human actions and they cannot forget that people give them significance, even when they are unwanted or unprepared (p.147).

Now it is possible to design analytical methods capable of producing cartographies that are fairly objective regarding the reality of, among others, social, communicative, economic, emotional, artistic, and political processes. Those analytical methodologies based on big data should be guided by scientific and ethical rigor in such a way that, as they go about enhancing our knowledge, they also provide protection for the subjects themselves that are generating such data without, in most cases, even being aware that they are doing so.

3.4. AI can enhance the quality of scientific publications

If with regard to university training programs it is essential to make a huge effort to achieve multidisciplinary coordination and creativity in order to offer the Bachelor's and Master's degrees required, there is another task that is even more urgent: that of guaranteeing the highest quality and innovation in scientific publications in general, and in the fields of Social Sciences and Humanities in particular. And this should be done by unequivocally adopting the advantages that are already to be had from big data and AI. The inalienable objective that Data Science makes available to publications and researchers in general is to incorporate methods for evaluating the effective content of what they publish for quality and, in the process, detect whether what is published offers real innovation. By and large, current methods that assess the quality of journals and editorials are based on formal qualitative and quantitative aspects, which have been used to rate the scientific production of the country, research bodies and even the researchers themselves. To date, the usefulness of such procedures and their ratings has not brought about agreement among the specialists on evaluating scientific documents nor does it satisfy many authors (Aguillo, 2015). Furthermore, the scant attention paid to publications on Social Sciences and Humanities has led to the creation of new tools and measures that also pose problems, since all researchers are aware that the privileged position of a specific journal in some rankings will in no way guarantee the quality of the real intellectual interest of the content published (López, 2016). This state of affairs is a huge obstacle that seriously hinders scientific progress.

With respect to measuring the impact of the publications on the scientific community internationally, the *Journal Citation Report*, known to all and currently published by *Clarivate Analytics*, or the *Scopus* database, which through the facilitation of abstracts and Website performance indicators, endeavors to improve and complete the results of the WOS (*Web of Science*).

Bibliometric works are based on the use and analysis of data from scientific publications and have given rise to interesting fields of research such as Webmetrics and *Alternative Metrics*, which are beginning to yield hopeful results, complementing those mentioned above.

The development of Internet and the Web, digitalization and open access, are generating a huge amount of data, whose processing and analysis using AI techniques facilitate more reliable information on the impact of scientific publications as well as generating new knowledge.

With regard to qualitative procedures, peer review often exposes shortcomings like the lack of clear criteria for the appraisal of proposals or the generous deadlines it demands.

The use of specific algorithms base don Cartesian genetic programming enables the review process to be reduced by about 30% (Mrowinski, Fronczak, Ausloos & Nedic, 2017).

One fundamental element that affects the quality of periodicals concerns management procedures. The emergence of platforms for publishers of scientific works in open access, such as *Frontiers*[1], is an example because it focuses both on providing assistance to editors, reviewers and authors, as well as the technical development of tools that make improving the quality of the scientific publications easier at all levels, including the identification of potential reviewers. To do so, it uses AIRA (*Artificial Intelligence Review Assistant*), which harnesses AI and machine-learning techniques and helps editors and reviewers, by facilitating their interaction and offering rapid results. This safeguards quality, rigor and independence, which presumably provides a guarantee for the authors. AIRA is an instrument that is designed to assess the quality of the work and facilitate the identification of specialists to carry out the review. Its algorithms identify overlaps with other studies, plagiarism, and flag up the quality of language and other aspects. It is regulated by processes that need to be followed by those texts that satisfy the quality criteria, and others for those that do not meet such criteria. This all means there is a significant reduction in review times. Finally, the platform is generating new metrics that will make it possible to assess the impact of academic articles objectively, along with other resources such as *Loop*, a network of scientific researchers that facilitates the dissemination of published books and articles.

In the same vein, *ScholarOne* is a peer review platform, employed by the press, which uses UNSILO, a program that analyzes written manuscripts through natural language

processing and machine learning. This allows principle statements and other functions to be extracted, thereby reducing the text to a script. It also detects similar ideas in other works, allowing any instances of plagiarism to be located and facilitating the localization of texts on similar topics. The application *Wizdom*.ai, which facilitates collaborative writing, is compatible with *Google Docs*, and uses databases, enabling connections to be established between disciplines and concepts, topic reviews, etc. Others like *Penelope.ai* examine the structure and citations of a text, as well as the statistical studies it contains. With this last function, *StatReviewer* is also used. These are tools that represent a huge step forward in offering contributions of much higher quality and also greatly reducing the amount of work required at the review stage. Since, in the latter case, their algorithm is capable of conducting a numerical assessment of a work.

Better known is the case of *Open Journal System*, which is starting to be used at universities and research centers both nationally and internationally, facilitating the various different management processes for such publications.

AI is also useful for performing repetitive tasks in journal editing, such as providing authoring assistance, storage, description and indexation in repositories, review by personnel not working for the journal, and dissemination on social networks, which can be coupled and decoupled in a systematic way (Priem & Hemminger, 2012).

Techniques like Natural Language Processing (NLP) and Machine Learning (ML) have not stopped developing and continue to do so today, being applied to all types of text, including scientific ones. The extraction of descriptive data and other metadata can facilitate rapid consultation and provide the most significant ideas.

In addition, the implementation of AI-based software, making it possible to describe both sound and visual documents by scanning them, can be of fundamental assistance in optimizing retrieval and dissemination of these types of documents, which very often accompany further documentation.

The new formats of scientific studies published online allow for the rapid, mass processing of data and its incorporation in author profiles and publication components. The incorporation and display of the data is immediate. Journal-publishing giants such as

Elsevier are implementing such technologies, which also need to become better known and adopted by more modest publishers so as to continue with their dissemination work.

Such technologies enable a huge amount of usage data to be extracted from both documents and logs, which has created a new analytical horizon for these publications (Jamali, Nicholas & Huntington, 2005).

4. Conclusions

In this study we have addressed three issues that might be considered of strategic importance for tackling professional and scientific adaptation to the possibilities and demands posed by progressive datification and subsequent processing using AI. The approach adopted in this research sets the following three matters in perspective: 1) the need to update the set of digital skills for the efficient analysis of big data as a base for drawing up the professional profile of the cyberanalyst.; 2) the assumption that AI is offering new epistemological opportunities in Social Sciences and Humanities that should be taken advantage of, and 3) the implementation of procedures deriving from AI for the effective analysis of the content of scientific publications when assessing quality and innovation.

These three issues have a direct relationship with the training of skilled personnel. The implementation of points 2 and 3 should contribute to generating an appropriate theoretical and methodological framework within the academic world that can facilitate the training that companies are demanding. In this regard, Education and scientific research will need to demonstrate a vision for the future and not wait for problems to worsen before they react. Ideally, advantage should be taken of the fact that, even in such a fast-moving environment, there is still time for a transition to come about that is neither traumatic today nor unsustainable in the future.

Finally, it is neither easy nor advisable to make predictions about how things will play out. The most sensible thing to do is to design potential strategies for improvement based on what is today already verifiable and what is presumed to be desirable. Attentive observation of the evolution of AI should be the best guide, not to make forecasts, but to make the most of opportunities to benefit society as a whole.

Bibliography

Abril, G. (2005). *Teoría general de la Información. Datos, relatos y ritos*. Madrid: Cátedra.

Aguillo, Isidro F. (2015). La Declaración de San Francisco (DORA) y la mala bibliometría. *Anuario ThinkEPI*, v. 9, pp. 183-188.

Ala-Mutka, K. (2011). *Mapping digital competence: Towards a conceptual understanding*. Seville: JRC-IPTS/European Comission.

Basu, A. (2013). Five pillars of prescriptive analytics success. *Analytics Magazine*, 8-12.

Bertalanffy, L., Ross Ashby, W., Weinberg, G.M. et al. (1978). *Tendencias en la teoría general de sistemas*. Madrid: Alianza Editorial.

Chen, H., Chiang, R. H. & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS quarterly*, *36*(4), 1165-1188.

Cleveland, W. S. (2001). Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review*. 21–26.

Damasio, A. (2018). *El extraño orden de las cosas. La vida, los sentimientos y la creación de las culturas*. Barcelona: Planeta.

De Mauro, A., Greco, M., Grimaldi, M. & Ritala, P. (2017). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*. *54*(5), pp. 807–817.

Digital paradigm. (2019). En *Dictionary Search*. IGI Global: Hershey, Pennsylvania. Retrieved from : https://www.igi-global.com/dictionary/digital-paradigm/7680.

European Parliament and the Council. (2006). Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. *Official Journal of the European Union*, L394/310.

Ferrari, A. (2013). *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*. Seville: JRC-IPTS. Retrieved from http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=6359.

Freund, J. (1975). *Las teorías de las ciencias humanas*. Barcelona: Ediciones Península.

Giner, S. (1974). *Sociología*. Barcelona: Ediciones Península.

Granville, V. (2014). *Developing analytic talent: Becoming a data scientist*. John Wiley & Sons.

Heeks, R. (2009). The ICT4D 2.0 Manifesto: Where Next for ICTs and International Development? *Development Informatics Working Paper* no.42 , IDPM, University of Manchester, UK. Retrieved from http://www.gdi.manchester.ac.uk/

research/publications/other

Heeks, R. (2016). Examining "Digital Development": The Shape of Things to Come? *Development informatics. Working paper series*, 64. University of Manchester. Retrieved from http://www.gdi.manchester.ac.uk/research/publications/other

Hunyadi, M. (2015). *La tiranía de los modos de vida. Sobre la paradoja moral de nuestro tiempo*. Madrid: Cátedra.

International Council for Science: Committee on Data for Science and Technology. (2012, April). CODATA, The Committee on Data for Science and Technology. Retrieved from International Council for Science : Committee on Data for Science and Technology: http://www.codata.org/

Jamali, H. R., Nicholas, D. & Huntington, P. (2005, Diciembre). The use and users of scholarly e-journals: a review of log analysis studies. En *Aslib Proceedings, 57*(6) pp. 554-571. Emerald Group Publishing Limited.

John W. T. (1962). The Future of Data Analysis *The Annals of Mathematical Statistics 33*(1), pp. 1-67. doi:10.1214/aoms/1177704711.

Latour, B. (1993). *Nunca hemos sido modernos*. Madrid: Editorial Debate.

Lope, V. (2018). La recuperación del sujeto en los datos masivos. En Lope, V., Marta-Lazo, C., Gabelas, J. A. (coords.) *Investigaciones en datificación de la era digital.* pp. 137-154. Seville: Egregius.

Lope V., Vidal Bordes F.J. & Mamaqi X. (2018). Datificación, big data e inteligencia artificial en la comunicación y economía. En Marta-Lazo, C. (coord.) *Calidad informativa en la era de digitalización: fundamentos profesionales vs infopolución.* pp. 65-82. Madrid: Dykinson.

López, W. (2016). Reflexiones sobre la medición de la calidad y el impacto de las revistas científicas. *Universitas Psychologica*, *15*(4).

Mamaqi, X., Miguel J. & Olave P. (2011). Evaluation of the importance of professional competences: The case of Spanish trainers, *On the horizon*, *19*(3), 174-187.

Mamaqi, X., Marta-Lazo, C. & Pérez R. (2019). Competencias y habilidades digitales en el mercado laboral europeo: el desarrollo de un marco conceptual para su medición. En Mancinas-Chávez R., Moya López D., *Comunicación emergente. Libro de resúmenes del IV Congreso Internacional Comunicación y Pensamiento*. Seville: Egregius.

Marr, R. (2016). How The Citizen Data Scientist Will Democratize Big Data – Forbes. Retrieved from http://www.forbes.com/sites/ bernardmarr/2016/04/01/how-the-citizen-data-scientist-willdemocratize-big-data/#32b0bf124557, Acceso enero 2018

Miller, S. (2014). Collaborative Approaches Needed to Close the Big Data Skills Gap. *Journal of Organization Design*, *3*(1), 26-30.

Mrowinski, M. J., Fronczak, P., Fronczak, A., Ausloos, M., & Nedic, O. (2017). Artificial intelligence in peer review: How can evolutionary computation support journal editors?. *PloS one, 12*(9), e0184711.

Naur, P. (1966). The science of datalogy. *Communications of the ACM*. *9*(7), 485-492.

Patil, D.J. & Mason, H. (2015). *Data Driven: Creating a Data Culture*. Sebastopol, CA: O'Reilly.

Priem, J. & Hemminger, B. H. (2012). Decoupling the scholarly journal. *Frontiers in computational neuroscience*, *6*(19), 1-13.

Río, O. del & Velázquez, T. (2005). Planificación de la investigación en Comunicación: fases del proceso. En Berganza, Rosa y Ruiz, José (coords.). *Investigar en Comunicación, Guía práctica de métodos y técnicas de investigación social en Comunicación*, pp. 43-76. Madrid: Mc Graw Hill.

Russom, P. (2011). *Big data analytics. TDWI Best Practices Report*. Fourth Quarter, 1-35.

Stephens-Davidowitz, S. (2019). *Todo el mundo miente. Lo que Internet y el "big data" pueden decirnos sobre nosotros mismos*. Madrid: Capitán Swing.

Tukey, J. W. (1977). Exploratory data analysis. *Reading: Addison-Wesley*.

Wajcman, J. (2017). *Esclavos del tiempo. Vidas aceleradas en la era del capitalismo digital*. Barcelona: Paidós.

White, M. (2012). Digital workplaces Vision and reality. *Business information review*, *29*(4), 205-214.

Zbigniev, H. P., Chand Seal, K., Leon, A. L. & Wiedenman, I. (2017). Skills and Competencies Required for Jobs in Business Analytics: A Content Analysis of Job Advertisements Using Text Mining. *International Journal of Business Intelligence Research*, *8*(1), 374-384.

Notes

[1] https://reports.frontiersin.org/