The Bible of AI™ OpenScience • Authors' publications

Data as the main focus of “State of the art of data science in Spanish language and its application in the field of Artificial Intelligence”

**Celia Medina Lloret, Adrián Hernández González, Diana Díaz Raboso,
Jose Alejandro Orozco Leal, Ángel Manuel Pérez López,
Marta Riquelme**

Published on: Jun 26, 2021

License: [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License \(CC-BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Abstract

This project is oriented to data science and artificial intelligence, involving data in Spanish language. It is based on our previous publication: *State of data science in Spanish language and its application in the field of Artificial Intelligence* <https://doi.org/10.21428/39829d0b.f5257ea7>. The main objective of this paper is split in two parts. Firstly, it is needed to verify if there exists a cultural bias for data science in Spanish language. Secondly, we have to check how privacy data is controlled in apps, which use contact tracing techniques, and also electronic devices such as smartwatches. To carry out the first part of the study, we have investigated about datasets in English and Spanish language and its technical structure. As well as, it has been designed a database, used for demonstrating the cultural bias that exists between English and Spanish language. For the second phase, it has been explored themes such as smartwatches for minors, adding a comparative table of privacy about important consumption brands. Following the assessment, it is observed that there exists a cultural bias between Spanish and English language. In fact, the results shows that a 70% of analysed datasets are written in English. The lack of inversion in technological education is one of the reasons why Spanish speakers countries lack of an appropriate technological education. On 14 May 2020, the newspaper The Economist writes an article about the involvement of countries on technological innovation. These countries do not include Spain because Spanish inversion does not cover more than 1,25% out of the PIB total. Europe has been driven by achieving the objective of reaching, at least, 3% in 2020. So, Spain is lower than the European media. On contrary, there are other countries such as South Korea, Denmark, and Sweden, which the barrier of 3% is reached and even 4% ([The Economist](#)). Consequently, it has negative effects for technological education for Spain. One proof of this is the results of the PISA index (Program for International Evaluation of Student), where Spanish obtain fewer qualifications in technology, science, and mathematics than the average of OCDE countries (Organization for economic Cooperation Development). Apart from Spain, Chile, México, and Columbia are also at the bottom of the list ([Epdata, 2020](#)). Moreover, this paper shows data that warn us about the importance of safeguard our security towards the technological breakthrough. Centralizing decisions in collaborative and international organizations by applying: efficient, ethics and deontological strategies could be a possible solution.

Key words: data science, artificial intelligence, Spanish language and cybersecurity.

1. Introduction

Artificial intelligence capacities are possible thanks to algorithms and data which train those algorithms. When observable and private or public data are not available, artificial intelligence does not exist. Definitely, artificial intelligence is data. ([BBVA, 2018](#)). If we work with data in artificial intelligence, it is important evaluating the quality of these data. According to Jordi Calvera Sagué, regional manager of InterSystems in Spain, Portugal, Israel, Greece, Turkey and Latin America, defends the idea of data quality when a project is created, for example, about machine learning.

Data scientists, who are persons dedicated to data science, affirm that data manipulation without an appropriate treatment, it has become in a great challenge for their job ([Calvera, 2020](#)). But, what do they mean with an appropriate treatment? This technique is known as Scrubbing. It is a technique of cleaning, whose aim is changing or eliminating incorrect, incomplete and duplicated data of a database ([TechTarget, 2019](#)).

In addition to the verification of the state of data, it is very important to preserve the integrity of used data, particularly if they are used to gather personal data. From 90's onwards, the concern about the lack of protection of information in network IT systems is justified by the increasement of command sequence. These commands are used to violate personal information of clients in a corporative, attacks of civil infrastructures and cyber espionage directed to secret documents ([Mulligan & Schneider, 2011](#)). Knowing these data, privacy and data protection is indispensable in this technological era, where cyber attacks are much more frequent, especially in large financial and political organisms ([Banafa, 2018](#)). That no means that any individual with an electronic device has not risks. For instance, if he does not take appropriate security measures, the privacy of his acts might be exposed.

2. Availability of datasets in Spanish and English language

This section discusses about datasets in Spanish and English language. The methodology consists of selecting search tools on Internet ([Stacy Stanford, 2018](#)): Dataset Search de Google ([Google, 2018](#)), Kaggle ([Goldbloom, 2010](#)) and World Bank Open Data ([BancoMundialdeDatos, 1944](#)). Then, some words related to artificial intelligence are filtered, thus indicating the availability of data in different languages, especially in Spanish language. To begin, they are analysed datasets related to linguistics, social networks, tourism, technology, and globalization.

Image 1

Firstly, the search tool, Dataset Search Google, has been selected as the basis of measurement. This graphic shows the predominance of English language in all sections, although it does not represent a big difference between values. However, there exists a significative difference of 200 datasets in the term “tourism” in English language. If we focus the attention on the word globalization in Spanish language, we find datasets related to globalization and international relations, infrastructures in the process of globalization, as well as frontiers and illegal markets in the globalization era. If one compares the data in percentage about languages, a 62.93% of datasets are written in English language and the rest 37.07% in Spanish. Although a large majority datasets come from Spain, it is also certain that a 20% of information come from Latin America countries such as: Mexico and Guatemala.

This schema is like the previous one, because the variables which are represented are the same. In this case, the search tool is Kaggle. However, the results given are dissimilar in comparison to Dataset Search Google. Once more, the language par excellence is English. Spanish language has not any graphic representation, obtaining values between 0 and 4 in each selection. As for the percentage, datasets found in Spanish language do not covers 2.2% out of total, meanwhile English language stands at around 98%. It also draws attention that any dataset of high social interest has not

been collected in Spanish language, for instance: globalization or social networks. On the contrary, English language has many datasets about mobile activity in the city or the falsity in face detection. Moreover, the date of updating of datasets found in English language is more recent than the Spanish language ones. It entails a greater involvement of English language with technology.

Next points are samples of articles which speak about science and technology in the World Bank Open Data, Dataset Search Google and Kaggle. Then, we compare and analyse the similarities and differences between them.

Image 3

This bar chart shows technological data in the World Open Data Bank to test the differences between English and Spanish language. In the scale of 0 – 100, English language covers 90 points, making a difference of 80 points with Spanish language. Considering the basis of technology, these data which are represented in the graphic make reference to the articles in scientific and technical publication in 2018.

The evaluations of Dataset Search Google are in line with the World Open Data Bank. In this case, it is not clear a disproportion of numbers but represented bar charts in Spanish and English language are very similar, existing a minimum difference. In

Spanish language selection, it is interesting to highlight some datasets, for example: the gender gap of graduated people in technological professional careers grouped by regions and the number of medical groups of high tech in Asturias by type. However, there is also datasets in Spanish language do not come from Spain but other countries such as Chile and Panamá.

Image 5

The previous graphic is like the World Open Data Bank one because it represents a big difference of data. It could say that Spanish language represents a 0% of information, meanwhile English language represents a 100%. This is another reason that digital world grows in leaps and bounds, and English language is a master key in this path.

The last two schemes are specific in artificial intelligence area.

The first one refers to Dataset Search Google. All selected themes have a major difference between results obtained in Spanish and English language. Despite of this, artificial intelligence is the most highlighted with 24 datasets. Presumably, this is because AI is a general area of technology, and it is constantly changing. In this research, it is found datasets related to classification of patrons in seismic images, applying artificial intelligence. These datasets also speak about the opportunities and

challenges of artificial intelligence and cybersecurity, the global sales of Natural Programming Language market 2017-2025 and second-hand cars for sale in Spain, which are managed by data science. As can be seen, there exist datasets of all Latin America nationalities, where multiculturalism and linguistic variety is noticeable. Despite of this, English language is the principal thrust of artificial intelligence. In fact, a 96.9% of datasets are written in English language, meanwhile Spanish language is represented by a 3.1%.

Image 7

Kaggle has low resources of datasets in Spanish and English language. In the case of Spanish language, it could be highlighted some datasets about artificial intelligence and data science (approximately a 1% found), meanwhile in English language all topics have a representation, covering until 100% out of results. Moreover, the number of datasets about data science in English language is the only data relevant in the graphic. In two datasets found in Spanish language, one comes from the Castilian variety in Latin America variety, concretely Paraguay. That is the same as occurs with the only dataset found in Spanish language about data science: organs donation in México, in which Spanish language of Spain is absent another time more.

As a conclusion, English language is the key in the development of datasets in any topics, especially in technology and artificial intelligence. Nevertheless, Spanish language still has a long way to go. For their part, Latin American varieties of Spanish language are filling the vacuum left by Spain in its relationship with technology.

3. Technical structure of datasets

We are conscious of the notable cultural bias that exist between Spanish and English language because of the table, located at the end of this section, about datasets in Kaggle. For this study, the methodology focuses on searching datasets in Spanish and English language, filtering terms such as:

- Artificial Intelligence
- Technology
- Machine learning
- Natural Language Programming
- Data science
- Data scientist
- Globalisation
- Linguistics
- Social networks
- Tourism

On April 12, 2021, they were found **23.771** datasets, of which only 10 "speaks" Spanish, representing a **0.042% out of total**. Meanwhile, **23.761** datasets "speak" English.

As regards the format of exportation of datasets, **CSV** format is the most used, concretely **14.931** datasets or **60.85% out of total**.

But what is the CSV format? It is a text format in which values are separated by comas. The reason why these files are useful to create datasets is because it is possible to convert them in a table in an automatic way. Moreover, these files require low space because they are flat file. The practice is that the file is separated by commas on each side, and it uses a line break to declare a new row. It can also occur that we find files separated by ";" or "-". In any case, it will be automatically converted in a table to open the file in a bearable software. Similarly, the image format more used are **JPG** and **PNG**, with **3.070 datasets (12.51% out of total)**.

Regarding databases, an **86%** are **static databases**, that means, databases which are

not updated since at least 90 days. On the contrary, there are **dynamic databases**, which are constantly updated (around **14%**).

Here below, the attached table is a repository that shows the values of each dataset. The datasets are detailed by language and exportation format.

Visit the web version of this article to view interactive content.

4. Software and tools in data science

At this point, it will be explored what programming languages are used in data science. During the revision, Scielo, Academia.edu, Google Academic, Kaggle and WorldWideScience are research sources used, what we get the majority of results in English language. By this way, it confirms the power of English language on data science.

After analysis of the results, we determine that programming language R is, par excellence, one of the most used in data science. According to the International Journal of Engineering and Technology Research (IRJET) ([Tecnología, 2020](#)), R is an open-source language, what implies providing a greater flexibility, security, and quality when it is offered a statistic analysis of data. In the same vein, R proportions many statistics and graphic capacities used for statistical evidence and classification. To this must be added public and private organizations that work with R Data Science: International Journal of Information and Education Technology, Kearney, Elaine de College of Arts and Sciences and Department of Computer Sciences ([Yamamoto et al., 2021](#)).

Another prominent programming language in data science is Python. It is a versatile language with a legible and clean code ([Robledano, 2019](#)). As with R, Python has an open-source that allows its application in any scenario. Furthermore, R has different libraries specialized in data manipulation, vector management, matrices, and mathematics operations rapidly, and the generation of interactive visualizations. All this controlled by a set of rules. The most outstanding ones are: Pandas, Numpy y Plotly ([Bagnato, 2020](#)).

According to Php doctor in Economy from University of Sao Paula, A. Días Porto Chiavegatto Filho:

"Programming languages R and Python go together." ([Chiavegatto Filho, 2015](#)).

This expression means that both programming languages have in common data management; recompilation, exploration, modelling and data visualization ([Parada, 2021](#)). The main difference is that R is oriented to statistical analysis of data science. Meanwhile, Python is multifunctional in other areas of web development, and it also allow integrating data which comes from different platforms ([Tecnología, 2020](#)).

Apart from these, there exists another powerful software in data science: Anaconda. It is one of the largest platforms of open-source that facilitates using other programming languages such as R and Python for data processing on a large-scale, predictive analyses and scientific computation. For this reason, it become in a useful tool in data science development ([Anaconda, 2021](#)).

With this in mind, it is very interesting to compile basic data about auxiliar tools used in data science.

Writing an inform in data science dispenses with mechanism with common denominators such as writing code and processing data. R Markdown, LaTeX or Jupyter notebook ([Randles et al., 2017](#)) are prominent software that can be used in data science ([Ding, 2016](#)).

Markdown was created by John Gruber in 2004 and distributed under the BSD's license. One of the advantages is its simplicity of syntax and easy access to mobile devices ([Cristóbal, 2016](#)). But why is Markdown an useful tool in data science? Firstly, Markdown offers collaboration with other data scientists and an environment to data science. It is not only used as a notebook for writing what you are doing but also what you are thinking. About collaboration, they can be used tools of controlling versions as Git and Github ([Wickham, 2016](#)).

According to Jupyter Project, it is an application of open-source, prominent for data transformation and visualization and machine learning. Likewise, it has other extensions as JupyterLab in which one of its data science connections consists of configuring and planning user interface to support works in data science ([Jupyter, 2021](#)). Jupyter Notebook, whose name comes from the three programming languages that supports: Julia, Python and R, it uses an ordinated list of input and output cell that it can have code, mathematics forms and text.

To conclude, it will talk about Sweve, a component of programming language R that allows working together with LaTeX. LaTeX is one of the typography programs capable

of producing complex mathematics equations ([Zhang, 2017](#)). In fact, during years it has been used to manage science, enginery, and mathematics journals ([Britannica, 2021](#)). It can be said that LaTeX is used to write scientific formulae in publications. It is important to add that LaTeX is very recognised in the world of science.

In next section, the focus is on protection measures of digital privacy applied in data management.

5. Cibersecurity in data

Currently, there exists a widespread availability of portable or "intelligent" devices such as digital cameras, agendas, tablets, and smartwatches. Recently, an article was published in the newspaper "ABC" on 11 may, 2017, about the lack of information of minors and the risks of Internet. In fact, many studies realized in the Proyecto Hombre Association have effectively shown a 90% of young from 10 and 14 years, they have a portable device. As well, the majority of them use it for chatting and disseminate images and videos in social networks, without knowing some basic privacy rules ([Setién, 2017](#)). This problem increases when adolescents and adults have a lack of knowledge about electronical devices functioning.

When we refer to "the intelligence" of a device, we cannot forget that something can apport benefit and solutions if something, previously, knows what it is happening. It is obvious that data are fundamental in this process. For more information, knowing smartwatches advantages. Among other functions, a smartwatch can measure arterial pressure and temperature, interacting with reconnaissance activities, controlling all our devices connected to Internet, notifying, and reading an email from your wrist without acceding to smartphone, sending voice messages, controlling your social networks, and taking calls ([Yañez, 2015](#)).

To know more about "smartwatches" and its relation to human activities surveillance, we take as a reference this investigation article: An Analysis of Human Activities Recognition using Smartwatches Dataset. It was published by the International Journal of Advanced Computer Science and Applications in December 2020 ([Karim et al., 2020](#)). Smartwatches have sensors which identify human behaviours patterns and machine learning techniques, the Bayes rule, data processing and k-nearest neighbours. These processes create a big volume of information with which getting accuracy in the results. These sensors are very useful for monitoring human healthy and offering services to patient. This is because they can measure many physical activities such as walking, cycling, running and walking up and down stairs. The

sensors what we are referring to are: Global Positioning System (GPS), Wireless Fidelity (Wi-Fi) and Near Field Communication (NFC). These technologies that are easily connected and disconnected, they offer valued information about our personal life. In other words, we are feeding the "intelligence" of device without any precaution.

GPS services are not only included in smartwatches, but they are frequently used in minors tracking device. However, there exists a huge lack of knowledge about this. The way of walks of our child does not only known by their parents, familiar or nearest friends, but technological enterprises can also manage this information. In fact, a security breach can cost a child abduction, which would imply a crime of gravity ([Judd, 2020](#)). In this approach, it is interesting to mention the Anna's case. It is a real fact that the engineer, Maik Morgenstern, narrated in the blog Internet of Things in November 2019. Anna is a young girl who lives with her parents in Lücklemberg district. During summer holidays, Anna usually goes with her grandparents to Norderney because her parents' job. Anna often takes some short walks. On first thought, it is a history told by any person, but not, this relevant information has been told by a smartwatch ([Morgenstern, 2019](#)).

5.1 Smartwatches in minors

We have chosen the Anna's case, but as Anna's, there are several cases. The most relevant security breaches of smartwatches are in applications and connections to servers, which store data rather than the own device. However, it is a serious issue because it is not necessary to access to physical device to benefit from vulnerabilities. The most common weak points are credentials encryption and encoding of communication between application and server which handles data. Another aspect is the low cost of device. Mainly, if we refer to devices that are used by minors, data privacy is very unprotected. Apart from this, no cases of study they includes the General Registry for the Protection of Personal Data, what means it would be a serious crime against private security. However, in trademarks such as Nokia ([Clausing, 2018](#)), Samsung and Huawei ([Clausing, 2018](#)), these problems are not very frequent because the connections are encrypted. In this way, data are completely encrypted and the attacks by rough force would be limited ([Stykas, 2019](#)).

From trademarks more used by minors are Carl Kids Watch, hellOO! Children's Smart watch, SMA-WATCH-M2 and GATOR Watch. The problem of Carl Kids Watch is in its application. The application fails in the implementation of certificates to HTTPS secure connections, so any certificate could be accepted. Similarly, they are used non encrypted connections and they are stored a file in SD card in which reveal the

password of an user account in a flat file. Moreover, the communication between user and server is realized throughout HTTP connection and clear text, including the register or user login in an account ([Clausing, 2018](#)).

The electronic register in hellOO! Children's Smart watch manages information that is not encrypted. This would mean that if someone can get access, attackers have total control of user data, for example, calls, messages and GPS position ([Henke, 2018](#)).

SMA-WATCH-M2 is another brands designed to "protect" minors by GPS. Precisely, Anna, the young girl who we talked in previous paragraphs, wears on her wrist. This device has a negative punctuation in cybersecurity because attackers can obtain the ubication and listen or manipulate confidential conversations. The vulnerability of security of this smartwatch is in the web portal because the encryption is no total and server does not verify the authentication token. It can be accessed to user identifier with all information related to device use and coordinates. Similarly, all this can be reproduced in manufacturer's application. Therefore, the manufacturer can have control of all smartwatch functionalities as if he were the legitim user ([Morgenstern, 2019](#)).

On the contrary, the devices of Gator Watch are accessible, using IDOR attacks to sensible user data and specific functions such as GPS real coordinates, bidirectional and unidirectional calls, and voice notes. It also recovers personal information like name, age, weight and hight. Moreover, the web page, which monitors devices, has a vulnerability related to proxy server's use. In a few words, it lets reviewing the request sent to web server and obtaining the total access to platform. But how can it access? The process consists of changing a few values of server petition as if we were administrators. In this way, we can have access to all user's data ([OZA, 2020](#)).

And we conclude that these "intelligent" devices can write our life exactly. For this reason, we cannot forget that we are the only responsible that our personal information is taken and used by others, wearing an "smart" device. It is certain that it depends on the security hole of each mark. Therefore, in next section, we include a comparative table which describes some trademarks of smartwatches and recommendations of security.

5.3 Comparative assessment of security

They have been analysed four trademarks: Samsung ([Foundation_Mozilla, 2020](#)), Apple ([Foundation_Mozilla, 2020](#)), Fitbit ([Foundation_Mozilla, 2020](#)), and Garmin ([Foundation_Mozilla, 2020](#)). These brands have been chosen on basis of sales to

Samsung, Apple y Garmin. The idea is controlling the highest sales and adding one with market models in the face of database as Fitbit.

The findings are that all marks collect data non anonym, using GPS and user's health.

However, Samsung does not collect online data from 13 years minors. On contrary, Apple, Fitbit and Garmin do not make distinction of age.

Another question is analysing if devices share data with third parties. Unfortunately, the answer is yes. Nowadays, these four trademarks share data with companies, although not all of them in the same way. The objective is analysing metrics and sharing results; however the user's data are only sold and hired by Samsung. On the contrary, Apple is the only company which does not share data with publicity or marketing aims.

SAMSUNG		APPLE	
Recopilación de Datos		Recopilación de Datos	
¿Se realiza de forma anonima?	No	¿Se realiza de forma anonima?	No
¿Se almacenan datos del GPS?	Si	¿Se almacenan datos del GPS?	Si
¿Datos de salud?	Si	¿Datos de salud?	Si
Información en línea de menores de 13 años	No	Información en línea de menores de 13 años	Si
Datos Compartidos		Datos Compartidos	
La información recopilada se comparte con terceros	Si	La información recopilada se comparte con terceros	Si
Los datos se comparten para análisis	Si	Los datos se comparten para análisis	No disponible
Los datos se comparten para publicidad y / o marketing de terceros	Si	Los datos se comparten para publicidad y / o marketing de terceros	No
Los datos se venden o alquilan a terceros.	Si	Los datos se venden o alquilan a terceros.	No
¿Los terceros estan autorizados a conocer la identidad del usuario?	Si	¿Los terceros estan autorizados a conocer la identidad del usuario?	Si
Seguridad		Seguridad	
Protección de cuenta mediante 2FA	No.	Protección de cuenta mediante 2FA	Si
¿Todos los datos son encriptados en su envío? (Smartwatch - Smartphone)	No, solo algunos	¿Todos los datos son encriptados en su envío? (Smartwatch - Smartphone)	Si
¿Todos los datos son encriptados en su envío? (Smartphone - Servidor)	Si	¿Todos los datos son encriptados en su envío? (Smartphone - Servidor)	Si
¿Se almacenan todos los datos encriptados?	No, solo algunos	¿Se almacenan todos los datos encriptados?	Si
¿Se avisa al usuario si se produce una brecha de seguridad?	No.	¿Se avisa al usuario si se produce una brecha de seguridad?	Si
¿Se realizan auditorias de seguridad en el producto?	No disponible	¿Se realizan auditorias de seguridad en el producto?	No disponible

Image 9

.3 How could brands of smartwatches be attacked?

When speaking about security, Fitbit and Garmin do not offer any kind of information,

except for Apple ([C. S. P. Program, 2020](#)) and Fitbit ([Program, 2020](#)) in the case of data encryption in the process of sending.

Garmin ([Program, 2020](#)) and Samsung ([Program, 2020](#)) encrypt partially data in the process of sending between smartwatch and smartphone. As well, Apple is the only one of the four brands analysed which stores encrypted data. Moreover, if a hole of security exists, the system alerts. However, all mentioned brands respect GDPR. Adding, the external security audits of products that should be carried out by enterprises are unknown.

In respect of cyberattacks, they have been analysed four typologies: man in the middle, phishing, SQL injection and DROWN attack.

- **Man in the middle:** In this cyberattacks, attacker finds client/user, in this case the smartwatch and server, mobile phone which receives data. The attacker can read, change, or insert data as he wishes. Samsung and Garmin are exposed to these attacks because they send data from the smartwatch to mobile phone without encryption. However, Apple and Fitbit send it totally encrypted.
- **Phishing:** This technique consists of deceiving user by entering his information in a web or app very similar to the original, but this web or app belongs to the attacker. By this way, he catches the data, and he can use them to have access to the web or real app. Samsung is vulnerable because it cannot be realised an 2FA authentication. However, Fitbit and Garmin have not any information. Meanwhile, Apple would be more protected because it provides of 2FA.
- **SQL injection:** This attack consists of using a vulnerability of system to enter malicious SQL sentences in the affected database. Applying SQL there are different kind of attacks by getting access to the database tables. In past, Samsung and Apple had got problems with SQL injection and the servers which stored its data. For this reason, it could be possible finding future security problems. Garmin and Fitbit have not relevant information about this issue.
- **DROWN attack:** This vulnerability affects to SSLv2 protocol. As well, it lets decrypting secure communications by using TLS protocol between client and server. All the chosen brands are protected against these types of attacks. It is possible because its servers do not support SSLv2, which let decrypting TLS connections between client and server.

SAMSUNG		APPLE	
Formas Ataque		Formas Ataque	
Man in the middle	Si (Sólo de Smartwach a Smartphone).	Man in the middle	No
Phishing (Relacionado con 2FA)	Si	Phishing (Relacionado con 2FA)	No
Inyeccion SQL	Posible	Inyeccion SQL	Posible
Ataque DROWN	No	Ataque DROWN	No
¿Posibilidad de ataque a 3º con nuestros datos?	Si	¿Posibilidad de ataque a 3º con nuestros datos?	Si

FITBIT		GARMIN	
Formas Ataque		Formas Ataque	
Man in the middle	No	Man in the middle	Si (Sólo de Smartwach a Smartphone).
Phishing (Relacionado con 2FA)	No disponible	Phishing (Relacionado con 2FA)	No disponible
Inyeccion SQL	No disponible	Man in the middle	No disponible
Ataque DROWN	No	Ataque DROWN	No
¿Posibilidad de ataque a 3º con nuestros datos?	No disponible	¿Posibilidad de ataque a 3º con nuestros datos?	No disponible

Image 11

.4 Privacy standards

"The most important resource of the world is not oil, but data". This information is published by the weekly publication, The Economist ([Economist, 2017](#)). Its aim is alerting by power data to population. Moreover, when we talk about artificial intelligence, we only speak of algorithms and data. That means artificial intelligence learns data and then, it offers solutions and recommendations. That is the reason why data security should be regulated to avoid serious consequences. However, it is not only for enterprises and company, but also for people. According to the electronic engineer, Juan A. Lloret Egea:

"Cybersecurity is an area focused on protecting those that extort people from different ages and assaults their privacy, intimacy, money and freedom by using technological resources".

Cybersecurity attacks are more frequent. In fact, Juniper Research ([Research, 2021](#)), an entity which offers investigation services and technological sector analyses, emits a

recent inform that reports to the cost produced by security breaches in United States. The cost for global economy accounts to 4 million of dollars and 400 billion ([Villas, 2020](#)).

The higher authority in the world of data science is Openmined. For this reason, it is practically impossible to work on artificial intelligence and data without mentioning it. But what are the functionalities of Openmined? According to the oficial page of Openmined:

“Openmined is an open-source community whose objective is preserving the world privacy by decreasing the entry of private technologies of AI.”

Andrew Trask, the leader of Openmined, poses interesting questions about cybersecurity and data science in a blog. Those questions are related to sensitive data managing, for example, medical history, financial states, and private states. Andrew Trask, the leader of Openmined, poses interesting questions about cybersecurity and data science in a blog. Those questions are related to sensitive data managing, for example, medical history, financial states, and private states. Furthermore, Task talked about the big quantity of data that people generate with electronical devices. Data need to be protected and limited. For this reason, Openmined supports data privacy rules by developing a security area in data science ([Emma Bluemke & Kang, 2019](#)).

5.5 Spanish, European, Latine America, and international in data use

Personal data management should be regulated by a normative o privacy rule. To solve it, all geopolitics entities have an internal regulation. In this section, it will expose how data privacy are in European Union, United States and Latin America.

According to European Commission, the code exposes: what data are vulnerable, how data management is authorized, what regulation exists for minors and, what are the penalties for committing infringements. In Europa, data privacy is regulated by General Data Protection Regulation (GDPR), in which European Union stablishes a list of requirements for organizations to manage citizens personal data ([Reglamento, 2021](#)).

In this paragraph, there are included the most important aspects in European internal regulation. In the range of personal data are name and surname, internet protocol direction (IP) and medical data. According to European rules of data protection, an information is manipulable when: the person has given his approval, there exists a legal obligation, vital interest of that person is protected and there are themes for one

mission of public interest. Moreover, it is important to highlight that minors should be authorized for using social networks and downloading internet contents by their parents.

To finish with European ordinance, we refer to the management of infraction rules and sanctions from which it obtains this information:

The non-compliance of General Data Protection Regulation (GDPR) involves penalties of 20 million of euros or a 4% of the world business volume of an enterprise because of some infractions ([Comisión Europea, 2021](#)).

The European agreement is applied in all countries which are members of European Union. However, each country presents its own legislation. In the case of Spain, it has a national and official diary to manage the rules of privacy; the Official State Bulletin (BOE). BOE makes public the rule 3/2018, in 5 December, which deals with personal data protection and digital rights guarantee. The most of articles included in the BOE are listed in the GDPR, although there exist some changes. For example, the seventh article, which talks about minors personal data management, establishes the age of 14 years as limit of consent ([BOE, 2018](#)). Other European countries as Germany, it has developed its own privacy data rule called BDSG as an adaptation of RPDG ([Protección de datos en Alemania, 2018](#)).

Moreover, a penalty system has been created for those behaviours which do not respect the General Data Protection Regulation (GDPR) in relation to personal data privacy. This could imply a penalty of 50.000 euros and specific obligations of video surveillance, labour relations and the evaluation of profiles ([Alemania, RGPD, 2017](#)).

"Gap is a company prepared for the new rule of consumers privacy. It requires more transparency about how data are managed."

It is the caption of image described in The Wall Street Journal about Gap shop. According to the journal "La Vanguardia": the first official rule of privacy in United States comes into force in California at the beginning of 2020. CCPA (Consumer Privacy Rule) is like the General Data Protection Regulation (GDPR) ([Renter, 2020](#)).

This standard affects to the big technological entities such as Amazon, Google, Facebook, and Microsoft, as well as enterprises which have not physical presence in California, but they offer services to the state. According to Molins Renter, this standard outlines if an enterprise buys or sells data from 50.000 residents from the state of California for one year, the enterprise must redact an inform about what data

they manage and what they are doing with the information. Apart from that, the penalty will be applied if the economical incomes of an enterprise come from the sales of the vulnerable data from its clients. In fact, the CCPA could emit penalties of 7500 dollars because of violation of existing law. It was then that Facebook had problems. According to The Wall Street Journal:

"Facebook should pay a penalty of 5 billion of dollars. According to Protection's America Consumers, Facebook deceived users by managing recklessly their personal information.

Microsoft reported on applying this normative to all country.

The General Data Protection Regulation of European Union was pioneer in boosting the management of vulnerable information of citizens, respecting for their rights and duties. From that moment, some Latin America political institutions have to develop an ordinance to protect this question. The first country of adapting a regulation like the RGD was Brazil, with General Protecting Data regulation (LPGD) in 2018, came into force in February 2020 (["La Visión de América Latina Sobre El Reglamento General de Protección de Datos," 2020](#)). What is to the adopted law by Brazil? Despite of its basis on RGD, the LPGD has different features on data processing. The main tasks of the LPGD are the following: making studies by investigation entities, preserving the anonymity of personal data. Protecting physical security of interested and fulfil a legal obligation ([BOE, 2020](#)). In the same way, Argentine created its own Personal Data Protection Law (PDPA) ([LeyDatosArgentina, 2021](#)). The main objective is achieving a suitability between data transferences with European Union. Few Latin countries are regulated by this legislation, although Ecuador and Paraguay are undeveloped of a regulation for privacy data.

5.6 Privacy and citizens' rights in datasets

There are many search engines on internet that collect datasets about different interest topics: politics, economy, mode, medicine, music, cinema, astrology, gastronomy, technology, and education. Particularly, Google Dataset Search has around 25 million of datasets ([Heras, 2020](#)). In some cases, for instance, in health sector, there are data to identify people, that means, vulnerable data.

Ever since it began the crisis of COVID-19, it has been incremented the following of infected and their contacts. Moreover, a person infected by COVID-19 can have infected many people without knowing it. For this reason, the task becomes more complicated. This type of situations are frequent in a supermarket queue or when we

go shopping. From that moment, the investigators focus on technology, especially in smart devices. Smart devices can use sensors such as GPS, Wi-Fi, and Bluetooth ([Ahmed et al., 2020](#)) for locating other near contacts. It is because contact tracing is developed.

"The technique of contact tracing consists on finding those persons who are not reported as infected by investigating to which has been infected from a positive case" ([Leonie Reichert*, 2020](#)).

Having an application based on contact tracing, it can help to control the expansion of COVID-19. However, what happens with user privacy? According to IEEE, the architecture of the system is linked to the way of recollecting and managing received data. In apps which using contact tracing, they are distinguished three architectures: centralized, decentralized and hybrid (a mixture of centralized and decentralized architecture). The added graphics are reference from A Survey of COVID-19 Contact Tracing Apps, published by IEEE Access in July 2020 ([Ahmed et al., 2020](#)).

According to the technological publication, Xataka, a centralized system identifies users at individual level by using a central server which the health authorities control all received information ([Fernández, 2020](#)).

This figure describes the internal functioning of contract tracing in an application. The design of a centralized architecture is based on BlueTrace. BlueTrace is a protocol which preserves privacy in the process of contact tracing by using Bluetooth connection because of global interoperability. BlueTrace has been created for the register of decentralized proximity. As well as it complements the coordination of centralized contact tracing by health authorities ([Jason Bay, 2020](#)).

Image 12

The graphic is divided into eight stages. Firstly, an user should download the application and register his data: name, telephone number, range of age and postal code in the server. Furthermore, the server verifies the telephone number by sending a SMS with an Once Time Password (step 1). For this reason, the server registers a TempID, which is encrypted, and it is only valid for 15 minutes. The TempID and expiry time are transferred to user application (step 2). After a client having contact with other user application, the message is exchanged by using Bluetooth. This message does not reveal any kind of user private information because of TempID code verification (step 3). Then, what happens when there is an infected person? The health authorities confirm if the infected has installed the APP, and if so, this person authorizes for registering his data on the server (steps 4 and 5). As well as, the iteration of the server by using encounter messages with approximation values (steps 6). By this way, the server has information to be proceed by health authorities (steps 7). In the last stage, medical centres, and hospitals alert to the server to communicate infected users about their exposure of COVID-19 (step 8).

According to the technological journalist, Enrique Pérez, the relation between users and infected users by COVID-19 in decentralized systems are private. That means that

health authorities do not personally identify users, although users have availability in basic functionalities of the application ([Pérez, 2020](#)).

Image 13

The decentralised architecture is based on private and automatic contact tracing protocol (PACT). As we can observe, this graphic is divided into eleven steps. Firstly, user should register in the app (step 1). The electronic devices generate seeds for creating pseudonyms and chirps for one minute in combination with the appointed time. By this way, the privacy will be preserved (step 2). These chirps are periodically exchanged with other devices of near contact (steps 3 and 4). When an individual infected by COVID-19 is diagnosed (step 5), the generated seeds can be uploaded to the server, if they are authorized by user's verification code (steps 6, 7 and 8). In this way, the server is a medium for collecting what users are or not infected. In step 9, the server begins storing data which are received from users. After that, other different users can register in the App, download the first data of the server database (step 10) and do their own inspection (step 11) ([Ahmed et al., 2020](#)).

Now, we focus on the hybrid system. Hybrid system is a combination of the centralized and decentralized architecture. On the one hand, the management of TempID is controlled by devices to ensure the privacy of user. On the other hand, the analysis of risks and notifications are supervised by server. The objective of this system is dividing

the functionalities between server and devices to achieve a major efficiency and security in contact tracing ([Ahmed et al., 2020](#)).

In Europe, the idea was creating an unified application for preventing COVID-19. It consisted of a voluntary usage by using private Bluetooth. As well as, the privacy values of users and the compliance of the General Data Regulation Protection ([Fernández, 2020](#)). What does Spain say about this issue? The Spanish agency of protection data published in May 2020 a program which analysed the benefits and costs about technology usage for the fight against COVID-19: geolocation in social networks, geolocation of data collected by telecommunication operators, infrared cameras, chatbots, contact tracing applications and digital passports for immunity. It not only implies an alert for economy, but also for security. The reason of maintaining anonymity is not sure ([Datos, 2020](#)). Furthermore, Manuel Carro, director of IMDEA Software Institute of Madrid, pronounced about contact tracing applications:

"The bomb is that none model guarantees privacy at all. Privacy to what extent? There are things which are inherent to centralised or decentralised contact models ([Carro, 2020](#)).

We conclude that the technological achievement for controlling COVID-19 becomes dangerous if there is data maladministration.

5.7 Data private use: federated learning

To paraphrase Juan A. Lloret Egea, electric and electronic engineer, *there are governments and hospitals which control data to include them in datasets. The problem of using private data is that there could exist connections to identify people. For example, if we think about man who is around 30 years old, tall, brown hair and with thin complexity, we say the name of a person.* However, what are the initiatives to solve these problems with privacy? MIT Technology Review raised an issue about anonymous use of data ([Conner-Simons, 2020](#)). In 2017, Google offered a new perspective of machine learning, so that its algorithm learned from different distributed data sources in some devices. In 2017, McMahan & Ramage describes federated learning or FL as a general focus of "carrying code to data, rather than data to code" and dealing with fundamental problems of privacy, property and location data ([Bonawitz.et.al, 2019](#)).

The strategy of federated learning applies the access of artificial intelligence in the medical sector. It can be a good point to maintain a balance between information administration and patient's privacy. In fact, according to Ramesh Raskar, professor of

Massachusetts Technological Institute and director of Culture Camera Investigation of MIT Media Lab:

"There is a false dichotomy between the privacy of patient's data and their use in society. People are not aware of the situation. In fact, it is easy to get privacy and usage at the same time" ([Hao, 2019](#)).

There exist many articles that corroborate the declarations of R. Raskar, for instance:

"Federate learning differentially private for cancer prediction" ([Begulier.et.al., 2021](#)).

This publication sumps up the results of achieving equanimity between prediction performance and privacy budget. For this reason, a supervised model is trained to predict cancer of mama from genomic data which are divided in two virtual centres. Moreover, "Federate learning to prediction keyboards of mobiles" ([Hard.et.al., 2019](#)) is another publication that shows the advantages of federate learning. In this model, federate learning can get more quality in datasets, achieving a major recuperation of prediction. Generally, federate learning offers users the opportunity to have a control of his data because the only transfers are algorithms and not data without processing ([Hao, 2019](#)).

6. Examples of APPs that use data, languages disponibility, and influence of technology in human behaviour

Artificial intelligence and Big Data are very important in the fight against Coronavirus, allowing great progresses in a short span of time. However, an inappropriate use of Technology can suppose a risk for population.

In the twentieth century, some famous thinkers, and writers such as George Orwell and Aldous Huxley, predicted topics that they are occurring now. Orwell in his fiction work, 1984, spoke about a society completely digitalized and controlled by technology. In the same way that Aldous Huxley in his utopic work, A brave new world, described an immersed society to happiness and advanced technology ([Ruiza, 2004](#)).

According to BBVA foundation, a 60% of spanish consider Internet as a essential piece of their lives, as contrasted with 18% in 2008. Furthermore, a 90% uses diary Internet, meanwhile 1 out of 3 persons are connected to net every day ([FundaciónBBVA, 2021](#)). For this reason, we make this question: Are we a completely digitalized society? And, what is more, technology domines our life, Are we going to allow someone also do it?

In 2018, the government of China promoted a kind of credit at local level. The Chinese social credit has a blacklist in which some of citizens are included. Firstly, the authorities collect data of the acts of citizens. Secondly, a punctuation is generated in basis on an algorithm. This algorithm is used for deciding if a citizen takes part of this blacklist. The blacklist privates from some rights to citizens for a long time. Moreover, some technological giants such as Alibaba joins to this initiative with Zhima Credit. It is another kind of credit used to reward the good actions of users. Gaining access to VIPs sales in airports and contacting loans at more favourable interests are some of the rewards. Zhima Credit not only collects information about where you are on Saturday's mornings but also it has access to your credit card and your debts. These acts allow citizens to take part of "exemplary citizens list" ([Financial, n.d.](#)).

China proposes a QR's code system to use at global level to reverse the pandemic. By this way, users introduce their data, and they receive a code from heath service in their smartphones. The codes are three colours; green (it symbolizes the free movement) and orange red (they symbolize that the person must be on quarantine for two weeks). For China, it is the best way to combat COVID-19, however, Hangzhou city plans a permanent version which allows assigning a punctuation based on clinical history, checks, and life habits ([Mundo, 2020](#)).

In any case, this severe populational control is usual for Chine. According to Growth from Knowledge, only an 8% of Chinese internauts do not give their data in exchange for a reward ([Statista, 2017](#)).

7. Spanish language in data

In this project, we have included data as well as Spanish and English languages connection. However, we cannot conclude without covering the state of Spanish language in the world: data and graphics.

Over the years, it is observed how English languages is one of the most important language in the world because it is a lingua franca. Europa Press ([Epdata, 2021](#)) makes public an information confirmed by Ethnologue Journal about Spanish is one of the most spoken language in the world, despite of not being declared as an official language in more than 20 countries.

Unlike Spanish language, English language has been established as an official language in more than 50 countries. It not only influences in English language expansion but also in investigation areas such as artificial intelligence. As well, it is interesting to know the evolution of natives' Spanish speakers in the rest of countries.

For this reason, it will be easy to identify if Spanish language is being pushed into the background. According to Cervantes institute:

Around a 7% of world population speaks Spanish language as a native language. These numbers have been increased over the past 8 years. That means Spanish language is not declining ([Epdata, 2021](#)). In fact, this graphic shows how Spanish speakers will grow in the future. According to the Cervantes institute, Spanish language overcomes the number of English and French native speakers.

Knowing that Spanish language is a powerful language, why is a cultural bias in such areas as technological education and investigation in artificial intelligence area? If there are more Spanish speakers in the world, why is the English language predominant in data science, machine learning, natural programming language, deep learning, neural networks, and cybersecurity? If Spanish language is the second language more spoken in the world, why do Latino Americans students have not access to a decent education in these topics? The lack of solutions carries out to investigate who are the responsible. The answers of these questions are regulated by the eight faults statement: lack of political leadership, lack of technological leadership, lack of conscientious leaders, lack of economical resources, lack of education tools, lack of leaders which are in strategic positions where important decisions are taken, lack of social consciousness and lack of interest.

8. General aspects

Considering that our society is moving towards to technological society, people who control technology often use personal data to non-ethical purposes and technology "speaks" English language, we have to consider seriously these questions: If we are not able to train ourselves in our mother tongue, who or what will take control of the decisions that technology has to make to continue advancing? Where do we go? Does it coincide with where we want to go?

9. Conclusions

Cybersecurity is an essential area in artificial intelligence. The reason is because it improves services and systems by proportioning stability and privacy. Moreover, data science analyses datasets with the aim of precisising a result, so data science becomes in an exact branch in AI. Furthermore, in basis on our previous project, we focus on the essential work tool for a data scientist: datasets. What cultural availability is between the major search engines that store million of datasets? To answer this

question, we have designed a database which has English and Spanish datasets. These datasets are grouped by title, theme, language, updating date and URL.

Visit the web version of this article to view interactive content.

Conjunto de datasets en inglés

Visit the web version of this article to view interactive content.

Conjunto de datos en español

The evidence of the results shows that there exists a cultural bias between English and Spanish language. To know more about objectives and methodology of these dataset, you have to click on: <https://github.com/Itvia/Los-datos-como-eje-principal-de-la-ciencia-de-datos-en-el-idioma-espa-ol-y-la-IA>

To conclude, it seems us interesting to mention the new statement about artificial intelligence which the European Commission presented on 22 April 2021. It was an unforgettable day for European Union because this initiative is an artificial intelligence deployment in Europe ([CENCENELEC, 2021](#)).

Citations

1. Economista, E. (2020). *La innovación tecnológica se atasca en España y despunta en Europa*. <https://www.eleconomista.es/ranking-innovacion-universidades/noticias/10537379/05/20/La-innovacion-tecnologica-se-atasca-en-Espana-y-despunta-en-Europa.html> ↵
2. Epdata. (2020). *España en la prueba PISA, en datos y gráficos*. <https://www.epdata.es/datos/espana-pisa-datos-graficos/484> ↵
3. BBVA. (2018). *¿Qué es la inteligencia artificial y cómo funciona?* <https://www.bbva.com/es/que-es-la-inteligencia-artificial/> ↵
4. Calvera, J. (2020). *Los datos son el combustible de la inteligencia artificial*. <https://www.computerworld.es/tendencias/los-datos-son-el-combustible-de-la-inteligencia-artificial> ↵
5. TechTarget, Colab. (2019). *Los datos son el combustible de la inteligencia artificial*. <https://searchdatacenter.techtarget.com/es/definicion/Scrubbing-o-depuracion-de-datos> ↵

6. Mulligan, D. K., & Schneider, F. B. (2011). Doctrine for Cybersecurity. *Daedalus*, 140(4), 70-92. https://doi.org/10.1162/daed_a_00116
<https://direct.mit.edu/daed/article/140/4/70/26918/Doctrine-for-Cybersecurity> =
7. Banafa, A. (2018). *Inteligencia Artificial en ciberseguridad: retos*.
<https://www.bbvaopenmind.com/tecnologia/inteligencia-artificial/inteligencia-artificial-en-ciberseguridad-retos/> =
8. Stacy Stanford, R. I. (2018). <https://sitiobigdata.com/2019/12/24/datasets-publicos-para-aprendizaje-automatico/#> =
9. Google. (2018). <https://datasetsearch.research.google.com/> =
10. Goldbloom, A. (2010). <https://www.kaggle.com/> =
11. BancoMundialdeDatos. (1944). <https://datos.bancomundial.org/> =
12. Tecnología, I. (2020). <https://www.unir.net/ingenieria/revista/r-vs-python/> =
13. Yamamoto, N., Ishida, A., Ogitsuka, K., Oishi, N., & Murakami, J. (2021). Development of Online Learning Material for Data Science Programming Using 3D Puzzle. *International Journal of Information and Education Technology*, 11(4), 154-163. <https://doi.org/10.18178/ijiet.2021.11.4.1505> =
14. Robledano, Á. (2019). <https://openwebinars.net/blog/que-es-python/> =
15. Bagnato, J. I. (2020). <https://blogs.solidq.com/es/poder-del-dato/10-librerias-python-para-data-science-y-m> =
16. Chiavegatto Filho, A. D. P. (2015). Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiologia e Serviços de Saúde*, 24(2), 325-332. <https://doi.org/10.5123/s1679-49742015000200015> =
17. Parada, M. (2021). <https://openwebinars.net/blog/r-vs-python-en-que-se-diferencian-y-en-que-se-parecen/> =
18. Tecnología, I. (2020). <https://www.unir.net/ingenieria/revista/r-vs-python/> =
19. Anaconda. (2021). <https://www.anaconda.com/products/individual> =
20. Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017, June). Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. 2017

- ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). <https://doi.org/10.1109/jcdl.2017.7991618>↵
21. Ding, I.-K. (2016). <https://www.quora.com/profile/I-Kang-Ding> ↵
22. Cristóbal, J. (2016). <https://markdown.es/> ↵
23. Wickham, H. (2016). *R for Data Science: Visualize, Model, Transform, Tidy, and Import Data*. OReilly Media. ↵
24. Jupyter. (2021). <https://jupyter.org/> ↵
25. Zhang, Y. T. (2017). Is LaTeX a useful skill for Data Science? In *Quora*. <https://www.quora.com/Is-LaTeX-a-useful-skill-for-Data-Science> ↵
26. Britannica. (2021). <https://www.britannica.com/technology/Java-computer-programming-language> ↵
27. Setién, M. (2017). https://www.abc.es/familia/educacion/abci-mayoria-menores-no-conscientes-riesgos-internet-201705111609_noticia.html ↵
28. Yañez, C. (2015). <https://www.ceac.es/blog/que-ventajas-o-desventajas-tienen-los-smartwatches> ↵
29. Karim, S., Aqil, S., & Mahmood, N. (2020). An Analysis of Human Activities Recognition using Smartwatches Dataset. *International Journal of Advanced Computer Science and Applications*, 11(12). <https://doi.org/10.14569/ijacsa.2020.0111241> <https://thesai.org/Publications/ViewPaper?Volume=11&Issue=12&Code=IJACSA&SerialNo=41> ↵
30. Judd, B. (2020). https://www.abc.net.au/news/2020-01-08/gps-trackers-for-children-what-do-you-need-to-know/11627208?utm_source=abc_news_web&utm_medium=content_shared&utm_content=link&utm_campaign ↵
31. Morgenstern, M. (2019). <https://www.iot-tests.org/2019/11/product-warning-chinese-childrens-watch-reveals-thousands-of-childrens-data/> ↵
32. Clausing, E. (2018). <https://www.iot-tests.org/2018/08/nokia-steel-hr-no-weaknesses-in-the-security-check/> ↵

33. Clausing, E. (2018). <https://www.iot-tests.org/2018/08/nokia-steel-hr-no-weaknesses-in-the-security-check/> ↵
34. Stykas, V. (2019). <https://www.iot-tests.org/2018/08/huawei-band-2-pro-fit-and-secure/> ↵
35. Clausing, E. (2018). <https://www.iot-tests.org/2018/02/carl-kids-deceptive-security-promise/> ↵
36. Henke, D. (2018). <https://www.iot-tests.org/2018/02/helloo-childrens-smart-watch/> ↵
37. Morgenstern, M. (2019). <https://www.iot-tests.org/2019/11/product-warning-chinese-childrens-watch-reveals-thousands-of-childrens-data/> ↵
38. OZA, S. (2020). <https://spanning.com/blog/insecure-direct-object-reference-web-based-application-security-part-6/> ↵
39. Foundation_Mozilla. (2020). <https://foundation.mozilla.org/es/privacynotincluded/samsung-galaxy-watch-3/> ↵
40. Foundation_Mozilla. (2020). <https://foundation.mozilla.org/es/privacynotincluded/apple-watch-6/> ↵
41. Foundation_Mozilla. (2020). <https://foundation.mozilla.org/es/privacynotincluded/fitbit-versa-3/> ↵
42. Foundation_Mozilla. (2020). <https://foundation.mozilla.org/es/privacynotincluded/garmin-forerunner-series/> ↵
43. Program, C. S. P. (2020). <https://privacy.commonsense.org/privacy-report/Apple-Watch> ↵
44. Program, C. S. P. (2020). <https://privacy.commonsense.org/privacy-report/Fitbit> ↵
45. Program, C. S. P. (2020). <https://privacy.commonsense.org/privacy-report/Garmin-Vivofit-Jr.> ↵
46. Program, C. S. P. (2020). <https://privacy.commonsense.org/privacy-report/Samsung-Galaxy-Watch> ↵
47. Economist, T. (2017). <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data> ↵

48. Research, J. (2021). <https://www.juniperresearch.com/home> ↵
49. Villas, M. (2020). <https://sherpa.ai/blog/es/la-importancia-de-cyber-security-data-science/> ↵
50. Emma Bluemke, A. T., Antonio Lopardo, & Kang, N. (2019). ↵
51. Reglamento. (2021).

<https://truben.no/latex/bibtex/>

PREVIEW

- ↵
52. Comisión Europea. (2021). https://ec.europa.eu/info/index_es ↵
53. BOE. (2018). <https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673> ↵
54. Protección de datos en Alemania. (2018). <https://leyderecho.org/proteccion-de-datos-en-alemania/> ↵
55. Alemania_RGPD. (2017). <https://www.lexology.com/library/detail.aspx?g=e9d3d53c-dd6b-4a81-a543-64bf221ee745> ↵
56. Renter, A. M. (2020). <https://www.lavanguardia.com/vida/20200105/472713380363/california-estados-unidos-privacidad-consumidor-e-commerce-comercio-electronico.html#:~:text=La%20CCPA%20establece%20multas%20de,hizo%20en%20la%20Uni%C3%B3n%20Europea>. ↵
57. La Visión de América Latina sobre el Reglamento General de Protección de Datos. (2020). *Comentario Internacional*, 99-112. <https://doi.org/10.32719/26312549.2019.19.4> <https://revistas.uasb.edu.ec/index.php/comentario/article/view/1546> ↵
58. BOE. (2020). <https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673> ↵
59. Ley Datos Argentina. (2021). <https://docs.microsoft.com/es-es/compliance/regulatory/offering-pdpa->

[argentina#:~:text=De%20acuerdo%20con%20la%20Constituci%C3%B3n,de%20datos%20p%C3%ABlicas%20y%20privadas. =](#)

60. Heras, J. M. (2020). [https://www.iartificial.net/google-dataset-search-descubre-conjuntos-de-datos/ =](https://www.iartificial.net/google-dataset-search-descubre-conjuntos-de-datos/)

61. Ahmed, W., Bertsch, P. M., Bivins, A., Bibby, K., Farkas, K., Gathercole, A., Haramoto, E., Gyawali, P., Korajkic, A., McMinn, B. R., Mueller, J. F., Simpson, S. L., Smith, W. J. M., Symonds, E. M., Thomas, K. V., Verhagen, R., & Kitajima, M. (2020). Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Science of The Total Environment*, 739, 139960.

<https://doi.org/10.1016/j.scitotenv.2020.139960>

[https://www.sciencedirect.com/science/article/pii/S004896972033480X =](https://www.sciencedirect.com/science/article/pii/S004896972033480X)

62. Leonie Reichert*, B. S. **, Samuel Brack*. (2020).

[https://eprint.iacr.org/2020/375.pdf =](https://eprint.iacr.org/2020/375.pdf)

63. Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., Seneviratne, A., Hu, W., Janicke, H., & Jha, S. K. (2020). A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*, 8, 134577–134601.

<https://doi.org/10.1109/access.2020.3010226>

[https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9144194 =](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9144194)

64. Fernández, S. (2020). [https://www.xatakamovil.com/aplicaciones/asi-app-que-quiere-gobierno-espana-para-frenar-coronavirus-uso-voluntario-unificada-toda-europa =](https://www.xatakamovil.com/aplicaciones/asi-app-que-quiere-gobierno-espana-para-frenar-coronavirus-uso-voluntario-unificada-toda-europa)

65. Jason Bay, T. A. Q., Joel Kek, Alvin Tan, Chai Sheng Hau, Lai Yongquan, Janice Tan. (2020). [https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf =](https://bluetrace.io/static/bluetrace_whitepaper-938063656596c104632def383eb33b3c.pdf)

66. Error [https://www.xataka.com/privacidad/descentralizado-vs-centralizado-que-dos-sistemas-rastreo-proximidad-diferentes-uso-nuestros-datos =](https://www.xataka.com/privacidad/descentralizado-vs-centralizado-que-dos-sistemas-rastreo-proximidad-diferentes-uso-nuestros-datos)

67. Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., Seneviratne, A., Hu, W., Janicke, H., & Jha, S. K. (2020). A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*, 8, 134577–134601.

[https://doi.org/10.1109/access.2020.3010226 =](https://doi.org/10.1109/access.2020.3010226)

68. Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., Seneviratne, A., Hu, W., Janicke, H., & Jha, S. K. (2020). A Survey of COVID-19 Contact Tracing Apps. *IEEE Access*, 8, 134577–134601.
<https://doi.org/10.1109/access.2020.3010226> –
69. Fernández, S. (2020). <https://www.xatakamovil.com/aplicaciones/asi-app-que-quiere-gobierno-espana-para-frenar-coronavirus-uso-voluntario-unificada-toda-europa> –
70. Datos, Agencia. Española. P. (2020). *EL USO DE LAS TECNOLOGÍAS EN LA LUCHA CONTRA EL COVID19. UN ANÁLISIS DE COSTES Y BENEFICIOS*.
<https://www.aepd.es/sites/default/files/2020-05/analisis-tecnologias-COVID19.pdf> –
71. Carro, M. (2020). <https://maldita.es/malditatecnologia/20200430/contact-tracing-rastreo-contactos-apps/> –
72. Conner-Simons, A. (2020). <https://news.mit.edu/2021/mit-launches-data-privacy-focused-initiative-fod-0420> –
73. Bonawitz, et.al, K. (2019). TOWARDS FEDERATED LEARNING AT SCALE: SYSTEM DESIGN. *ArXiv*, Vol2(1902.01046v2 [cs.LG]), 1.
<https://arxiv.org/pdf/1902.01046.pdf> –
74. Hao, K. (2019). <https://www.technologyreview.es/s/11017/aprendizaje-federado-la-nueva-arma-de-ia-para-asegurar-la-privacidad> –
75. Error <https://arxiv.org/pdf/2101.02997.pdf> –
76. Error <https://arxiv.org/pdf/1811.03604.pdf> –
77. Hao, K. (2019). <https://www.technologyreview.es/s/11017/aprendizaje-federado-la-nueva-arma-de-ia-para-asegurar-la-privacidad> –
78. Ruiza, E., M. ., Fernández, T. y Tamaro. (2004). *George Orwell*.
<https://www.biografiasyvidas.com/biografia/o/orwell.htm> –
79. FundaciónBBVA. (2021). *El proceso de digitalización de la sociedad española se consolida y se amplía en el marco de la pandemia con nuevos usos como el teletrabajo*. <https://www.fbbva.es/noticias/el-proceso-de-digitalizacion-de-la-sociedad-espanola-se-consolida-y-se-amplia-en-el-marco-de-la-pandemia-con-nuevos-usos-como-el-teletrabajo/> –

80. Financial, A. (n.d.). *Zhima Credit*.

<https://gw.alipayobjects.com/os/rmsportal/nGVZldPeZtcyBkbhojbi.pdf> ⇐

81. Mundo, B. N. (2020). *Covid-19: el sistema de códigos QR que China propone usar a nivel global para contener el coronavirus (y por qué causa polémica)*.

<https://www.bbc.com/mundo/noticias-55049848> ⇐

82. Statista. (2017). <https://es.statista.com/grafico/7860/venderias-tus-proprios-datos-personales-si-pudieras/> ⇐

83. Epdata. (2021). <https://www.epdata.es/datos/lengua-espanola-mundo-datos-graficos/513> ⇐

84. Epdata. (2021). <https://www.epdata.es/datos/lengua-espanola-mundo-datos-graficos/513> ⇐

85. CENCENELEC. (2021). *European Standards support the EU ambitions on Artificial Intelligence*. https://www.cencenelec.eu/News/Brief_News/Pages/TN-2021-020.aspx ⇐