Article in Press IJIMAI journal

Combating Misinformation and Polarization in the Corporate Sphere: Integrating Social, Technological and AI Strategies

Alberto Tejero[1], Galena Pisoni[2], Ziba Habibi Lashkari[1], Sergio Rios-Aguilar[1]*

[1] Universidad Politécnica de Madrid (UPM), Madrid (Spain)
[2] York St. John University, York (United Kingdom)

Received 2 February 2024 | Accepted 18 March 2024 | Early Access 22 March 2024

Abstract

In an era where misinformation and polarization present significant challenges, this research examines the root causes within social networks and assesses how corporations can use AI technologies for prompt detection. This research uses a dual approach: a "telephone game" with 225 participants from a Spanish university to study the spread of misinformation, and interviews with 15 experts from three French tech companies to investigate technological solutions. The findings indicate that almost one-third of participants inadvertently contribute to polarization, and around one-quarter propagated misinformation. The study also identifies the existing tools enhanced by AI and Machine Learning that effectively detect misinformation and polarization in corporate settings. This investigation provides crucial insights for practitioners to strengthen their strategies against misinformation and technical challenges and opportunities.

Keywords

Business Strategies, Information Systems, Misinformation, Polarization.

DOI: 10.9781/ijimai.2024.03.003

I. Introduction

Society expects online businesses to maintain ethical and socially responsible. Assisting companies and practitioners in making in making informed decisions amid the rampant and unchecked spread of misinformation is challenging [1] [2]. In the world of social media individuals consume and produce information, with a single post potentially reaching vast audiences rapidly. Moreover, discerning the trustworthiness of information is difficult for individuals, who often receive it from their network and lack knowledge of the original source [3] [4]. The study addresses the phenomena of misinformation and polarization and their impact on companies' online presence including their products or services. We also explore user tendencies in a controlled experiment.

Online social networks enable global access to and rapid dissemination of information [5]. The content on these networks influences users' decisions and opinions on issues [6]. For many users, social media has become the primary source of information [7]. Nowadays, instead of traditional media, people rely more on social networks for information and the spread of information [5].

The problem of misinformation has garnered significant attention, and been a topic of concern for a long time [8] [9] [10]. Also, there has been focus on clarifying the elements that contribute the misinformation spread, identifying the user groups who share it, and understanding how businesses can defend their brands against it.

* Corresponding author.

E-mail address: sergio.rios@upm.es

This research addresses these gaps in the social computing landscape by pursuing two objectives: firstly to identify the factors that result in business-related misinformation and polarization, and secondly, to develop an automated system for such misinformation targeting businesses.

According to Guess and Lyons [11], misinformation consists of false or misleading messages presented as informative content, such as elite communication, online messages, advertisements, or published articles. Therefore, misinformation can be defined as a claim that contradicts or distorts common understandings of verifiable facts. On the other hand, disinformation is a subset of misinformation spread with the intent to deceive, distinguishing it from misinformation that may be shared without any intention of mislead.

Fig. 1 and Fig. 2 illustrate an increased interest in misinformation and disinformation at social (Google trends) and scientific (Web of Science) levels in recent years. However, an analysis of the scientific articles indicates that this interest focuses on the intentional spread of information [12] [8] [13] [14].

Polarization has been linked to differences in policy, known as issue polarization. Nowadays, it encompasses a wider range of effects such as effective polarization and distrust towards opposing views [15] [16] [17].

Our beliefs -and environment- can bias our perceptions without our awareness which can result in unintentional misinformation. The bar chart shows information about a specific event, which can always have different interpretations, depending on the observer even if the data remains the same for everyone.

Please cite this article in press as: A. Tejero, G. Pisoni, Z. Habibi Lashkari, S. Rios-Aguilar. Combating Misinformation and Polarization in the Corporate Sphere: Integrating Social, Technological and AI Strategies, International Journal of Interactive Multimedia and Artificial Intelligence, (2024), http://dx.doi.org/10.9781/ijimai.2024.03.003

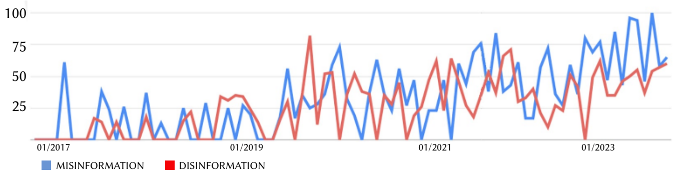


Fig 1. Analysis of the scientific publications of the concepts of Misinformation and Disinformation between the years 2017 to 2023 (Source: own elaboration based on Google Trends).

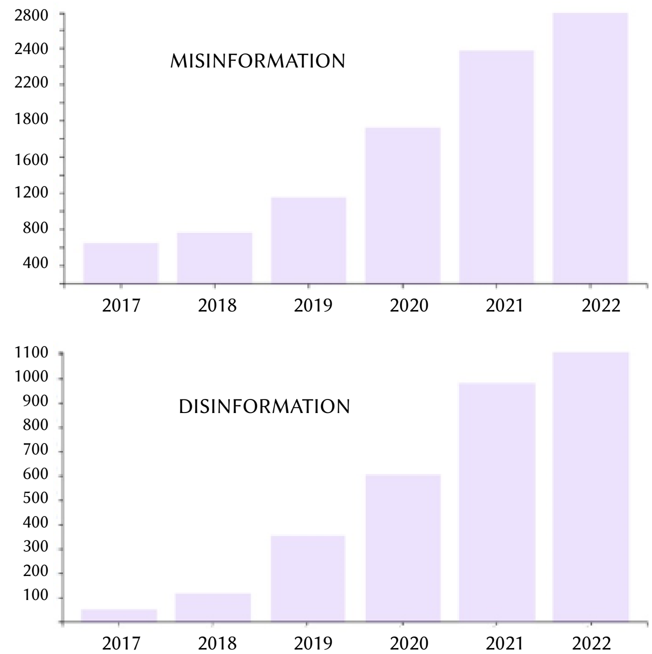


Fig. 2. Analysis of the scientific publications of the concepts of Misinformation and Disinformation between the years 2017 to 2022 (Source: own elaboration based on Web of Science).

This fact leads on many occasions to the generation of unintentional misinformation [16] [18]. That is why this study has been considered relevant in this work, to determine its level of importance and impact in relation to misinformation.

On the other hand, sociotechnical systems are intricate hybrids were human and technical resources collaborate to accomplish tasks [19] [20]. Performance depends on optimizing both the technical aspects -software and machinery infrastructure- and the social aspects such as, rules, procedures, roles and coordination.

Given the new demands of companies to constantly innovate and stay relevant, research in the domain of information systems (IS) concludes that IS are incomplete, always requiring enhancements. [21] [22]. As a result, the development of new technologies requires that sociotechnical constructions are revisited, and this drives the improvements developments of the IS of companies. These changes in the operations introduce new socio-technical constructions within the company.

The current trend towards digital transformation requires that companies maintain flexible and agile tech architecture that evolve with customer and organizational needs. Thus, modernizing information systems and technology in infrastructure is essential for integrating diverse data and delivering digital services [23]. Therefore, a flexible underlying architecture is crucial and can accommodate integrations of new modules with different functions suited for the company´s needs [24].

Architecture must also be robust enough to enable the company to develop new functionalities, instead of fixing bugs in the current system [25]. To achieve this, the technology architecture and design should be highly modular and loosely coupled, and support independent development of components and modules [26] [27]. The constantly evolving technological, organizational, and external environment means that business intelligence is being taken seriously in companies. Generally, these new services and component engineering cannot be provided without the insights gained through the application of Data Science and Artificial Intelligence (AI).

Artificial Intelligence has revolutionized how businesses utilize and manage data with Machine Learning (ML) -a subset of AI-, and automation, becoming ubiquitous in organizational operations. In recent years, AI has significantly helped companies in analysing and understanding their internal workings. AI models are on the rise, and while AI has contributed to misinformation [28], the social media literature has begun to investigate how AI can be used to predict and detect polarization and misinformation on social media [1] [29]. Still, literature has paid little attention to explaining how these models can be integrated into existing enterprise architectures, and how companies can leverage such available models. Another pressing issue is the lack of a clear definition of effective metrics for detecting misinformation and polarization, and what it means for a comment or opinion to be classified as polarized or misinformation [30]. Model accuracy, as applied as a metric in other fields, might not be relevant in this context. Despite its potential relevance to businesses, there appears to be a lack of research in this specific area to date.

Among others, Meta (formerly Facebook) is one of the companies that has utilized AI tools to identify and mitigate the spread of misinformation on their platform, implementing a range of policies and products [31]. These measures include adding warnings to content, reducing the distribution of questionable content, and removing harmful misinformation. They designed a specialized AI system to flag potential issues for review and to automatically detect new instances of previously identified misinformation, subsequently sending them to independent fact-checkers. They have developed technologies such as SimSearchNet++, which enhances image matching to identify variations of known misinformation images with high precision. Additionally, they have introduced new AI systems capable of detecting new variations of content that have already been debunked by fact-checkers, utilizing technologies like ObjectDNA and LASER cross-language sentence-level embedding.

This study investigates the phenomenon of users' information sharing about companies (or their products/services) on online social networking websites, and its implications for companies´ technological developments. This research examines the factors that affects individuals' knowledge-sharing behaviours in social settings.

In this work, we face a set of research questions:

- What are the factors that lead to the polarization and degradation of messages, unintentionally generating misinformation? What are the typical user profiles that polarize and generate this misinformation?
- How companies can respond to misinformation and polarization presence regarding their products and services on social media? How can companies design / integrate components that can help them detect polarization and misinformation? How can companies integrate them with their existing information systems and enterprise architectures?

The gap that we try to fill in with our research is related to the analysis and identification of misinformation and polarization factors. Specifically, as previously anticipated, the focus of this work is on the analysis and identification of the factors related to non-deliberate misinformation.

We believe the finding will offer insights for enhancing corporate online strategies, leveraging AI-based tools for automation, and strengthening user trust- a vital aspect for sectors such as banking and healthcare.

We design a test to facilitate user knowledge exchange, employing a socio-technical approach. This allows us to explore human factors influencing misinformation and polarization, like gender and age, as well as the retention of information in friend circle. We also examine how companies can adapt there IS architectures to address these issues on social networks.

A mixed-method approach was employed, combining computational social science techniques with developer interviews and case studies. This aimed clarify the factors diving message polarization and degradation, intentional misinformation about companies, the profiles of those who generate such content, and how companies can technologically detect and prevent this spread on social media.

The research findings and contributions are further discussed from both aspects, sociological and technical. We propose a way forward for corporate (digital) social responsibility, and how they can implement responsible social presence on online platforms.

The structure of this document is as follows: section II outlines the research questions and objectives, section III describing materials and methods, section IV detailing the results, Section V discussing the results and Section VI presenting the conclusions.

II. Materials and Methods

To explore our research topic, we will do experiments from two dimensions: technological and social.

This dual approach addresses gaps identified in the literature [14]: firstly, the factors which cause unintentional misinformation spread, and secondly, the absence of defined technological strategies for companies to manage online misinformation.

A. First Experiment: Social Dimension

The first experiment of this research was conducted at the empirical level within the social dimension, using the "Telephone game" method. Known globally by various names, including Chinese whispers, this game begins with one player whispering a sentence or phrase to the next person. This message is secretly passed along from person to person. When the secret has reached the entire group, the last person to hear it announces the secret aloud for everyone to hear. In general, by the game´s end, participants are surprised to hear how different the final version is from the version they heard. One of the main reasons for changes in the message is attributed to unintentional changes, such as impatience, wrong or faulty connections, although it can also be due to deliberate alterations on the part of the participants. The telephone game serves as a metaphor for the distortion that happens when information is passed along from person to person, whether first, second, third hand, or even more [32] [33] [34].

Specifically, this experiment aims to understand the factors causing misinformation and polarization through observation or direct experience with end users. Therefore, the focus of this experiment was on the users and their behaviour concerning misinformation and polarization, rather than the technology. For this reason, a low-tech test was designed, but one that would allow obtaining the maximum information about the users, who were the object of this phase of the research.

The designed test is described below (refer to Fig. 3 for more details):

- A text is selected, long enough (to offer the possibility of its transformation, summary, etc., at different levels), within a specific domain (news, politics, health, etc.).
- The original text is sent to a number N of people.
- Once they receive the text, each person must rewrite the text in their own words, creating a new version.
- After rewriting the text, each person must send the new text to a new group of people (M).
- After, each person must complete an anonymous form providing demographic information (age, gender, etc), which aids in group identification.
- People who receive the message must repeat steps 3 and 4 of the process to reach other people.
- The experiment concludes once a statistically sufficient sample is reached.
- Finally, the collected results obtained are analysed and categorized.

This experiment was designed using a simple technique, with the aim of facilitating participation, seeking to simplify the process as much as possible. In this way, the goal was to maximize the number of potential participants in the experiment in order to obtain a sufficiently significant sample for the research.

In this research, the topic of Covid-19 vaccines was used due to its significant global impact, making it a subject likely to draw heightened attention from participants. This can facilitate the identification of factors like misinformation and polarization. It´s important to note that the Covid-19 vaccine topic was not central to our research but was merely a means to conduct the study. The selected text comes from a real text identified on the social network X (formerly known as Twitter):

"*The SAGE group indicated that no serious allergic reactions caused by the AstraZeneca vaccine have been recorded in the clinical trials against coronavirus. However, as for all vaccines, this should be administered under medical supervision, with appropriate medical treatment available in case of allergic reactions. In addition, anyone with an acute fever (body temperature over 38°C) should postpone vaccination until they are afebrile. However, the presence of a minor infection, such as a cold or low fever, should not delay vaccination*".

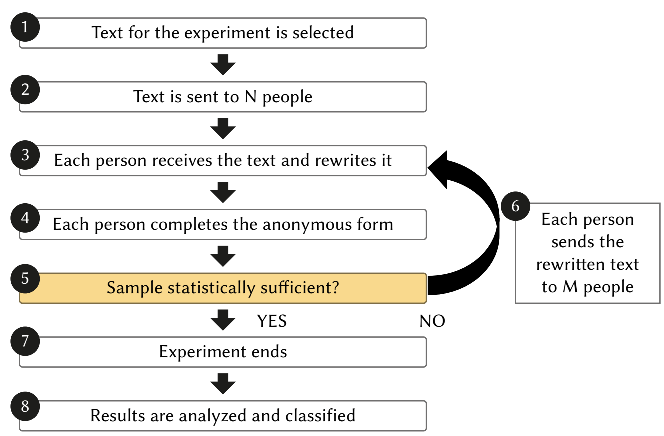


Fig. 3. Flowchart of the test designed for the experiment (Source: own elaboration).

The experiment was conducted within a university setting, encompassing students, faculty, and their close relatives, to ensure a diverse sample. To broaden the experiment´s scope and access the cultural dimension´s influence without adding undue complexity, it was executed in both English and Spanish.

To study misinformation, we identified specific keywords in the text to see how many keywords each user was able to retrieve and rewrite. The goal is to provide a clear view of how information is deteriorating regarding missing key values or, in this case, missing keywords.

For the polarization study, sentiment analysis was performed on each rewritten text, to observe how the polarity affects each user depending on the initial value of the original text. If the original text was identified with a positive voice, users who returned a negative polarity were analysed (and vice versa). Like this, it can be seen if people tend to have a positive or negative polarity based on their characteristics.

Equation (1) is used to obtain a statistically valid sample where N= Population size, z= Critical value of the normal distribution at the required confidence level, p= Sample proportion, e= Margin of error

$$Sample\ Size = \frac{\frac{z^2 * p(1-p)}{e^2}}{1 + (\frac{z^2 * p(1-p)}{e^2 N})} \tag{1}$$

The original text was sent to 96 people (N) in English and Spanish, each of whom had to send the original text to three other people (M) to complete the experiment. In the end, 225 users participated in this experiment. In particular, 142 completed the experiment in Spanish and 73 in English. The total time it took to conduct the experiment, from when the first messages were sent to when the last responses were received, was two weeks.

In the experiment, N=96 people were initially surveyed (69 in Spanish and 27 in English). Each of the people in turn had to get M=3 other people to complete the experiment. To statistically guarantee the reliability of the samples, the following was done:

In Spanish:

- 276 (69 + 69 x 3) people is the maximum we can reach (size of the universe), while the people who have actually responded to the survey (sample) were: 142 people.
- When performing the calculations using (1), the value 136 is obtained. Therefore, 136 surveys were needed to ensure the reliability of the sample.
- Given that 142 people responded to the survey (142 > 136), the study was statistically feasible, with a 95% confidence level and a 6% margin of error.

In English:

- 108 (27 + 27 x 3) people is the maximum we can reach (size of the universe), while the people who have actually responded to the survey (sample) were: 73 people.
- The value 70 was obtained when performing the calculations with (1). Therefore, 70 surveys were needed to ensure the reliability of the sample.
- Given that 73 people responded to the survey (73 > 70), the study was statistically feasible, with a 95% confidence level and a 7% margin of error.

B. Second Experiment: Technological Dimension

This part of research focuses on the technological aspects and how companies can detect misinformation and polarization related to their products and services. We also explore how they can design or integrate technological components to analyse data from social networks and identify such issues.

The goal of this research segment is to assist companies in detecting disinformation in communications, as well as the factors that produce polarization, positive or negative, or misinformation (how, when, why, by whom). Analysing real time information will enable online businesses to more effectively tackle misinformation and polarization on social platforms, allowing them to monitor and mitigate the factors that lead to disinformation and polarization in user message.

The interview guide was developed based on the conceptual frameworks [35] and [36]. Due to the exploratory nature of the study, the questionnaire consisted of open-ended questions. The interviews lasted between 30 minutes and 45 minutes. The open-ended questions guideline dealt with the following subjects:

- Current tech practices and the types of analysis companies perform on a daily basis regarding misinformation and polarization.
- How misinformation / polarization tools should be designed and chosen so to allow the online business and companies to be ethical, socially responsible, and protect customers.
- Integration of such solutions with the company´s existing systems.

We followed a qualitative process to determine the collection and use of customer data from social media with company representatives coming from different industries. First, we obtained relevant documentation from the companies on the use of social media data and tools for their analysis. We relied on the data sources as specified in Table I. A descriptive review approach was employed to review relevant documentation on the enterprise architecture used in online business and digital companies and to identify the typical patterns in social media data analysis. The data for the case studies was collected in the Spring of 2023 and through interviews, we gained insights into the companies' internal processes from the respondents 'perspectives.

TABLE I. Description of Companies Involved and Participants, With Their Identification Numbers Marked With R, in Technological Experiment (Source: Own Elaboration)

Company	Domain	People involved in the interviews
Case company 1	Technical platforms realizations	2 (Strategic project manager, R1, Technical manager, R2)
Case company 2	Implementation of big data solutions for companies	2 (Technical manager, R3, Project Manager, R4)
Case company 3	Digital business consulting and services company that develops new digital services for clients, the main tasks: consulting and service design	3 (Technical coach, R5, R&D manager, R6, Business Developer, R7)

III. Results

In this section we analyse the results we obtained from both experiments.

A. First Experiment Results

In the polarization analysis, the results obtained from the 215 participants of the experiment showed that 33.5% of the participants changed the polarization of the text they received. That is, approximately 1 in 3 people changed the polarization of the message they received without realizing it. In particular, in the experiment carried out in English, 37% of the participants changed the polarization of the text (27 out of 73), while in the experiment carried out in Spanish it was 31.7% (45 out of 142).

At the gender level, 34.71% of the female participants in the experiment changed the polarization of the original text they received, versus to 31.25% of the male participants. In particular, in the English version of the experiment, 38.3% of the female participants alerted the text´s polarization compared to 32.43% in the Spanish version. In the case of men, 34.62% alerted the text in English experiment, while 30.88% did so in the Spanish experiment.

At the age level, the details of the participants changed polarization can be seen in Table II.

During the misinformation analysis, a noteworthy 27% of the 215 participants revised their texts by incorporating half or less of the keywords present in the original material. That is, approximately 1 out of 4 people was unable to retain or maintain in their rewritten text more than half of the key words of the original text that they received. In particular, in the experiment carried out in English, 28.77% of the participants retained half or less of the keywords, while in the experiment in Spanish it was 26%.

TABLE II. Detail of Participants Who Changed Polarization at Age Level (Source: Own Elaboration)

Age range	Participants who changed polarization (%)	(English experiment) participants who changed the polarization	(Spanish experiment) participants who changed the polarization
18-30	35.16%	42.86%	30.38%
31-40	8.33%	14.29%	0%
41-50	61.54%	33.33%	14-29%
51-60	34.62%	25%	36.36%
+60	47.83%	0%	50%

At the gender level, 29.75% of women rewrote their texts using half or fewer of the original text´s keyword, compared to 23.4% of men. In particular, in the experiment carried out in English, 36.17% of the participating women retained half or fewer words, whereas in the Spanish language experiment, the figure was 25.67%. For men, the English experiment saw 15.38% retaining half or fewer keywords, while in the Spanish experiment, this was higher at 25.68%.

Finally, at the age level, the details of the participants who rewrote their text using half or less of the keywords of the original text (%) can be seen in Table III.

TABLE III. Detail of Participants Who Rewrote Their Text Using Half or Less of the Keywords of the Original Text (%) at Age Level (Source: Own Elaboration)

Age range	Participants who rewrote their texts using half or less of the keywords of the original text (%)	(English experiment) participants who rewrote their text using half or less of the keywords of the original text (%)	(Spanish experiment) participants who rewrote their text using half or less of the keywords of the original text (%)
18-30	24.22%	26.53%	22.78%
31-40	25%	28.57%	20%
41-50	23.08%	33.33%	14-29%
51-60	31%	50%	27.27%
+60	43.48%	0%	45.45%

B. Second Experiment Results

To implement new technological components and services for detecting misinformation and polarization, a company requires a holistic approach. This includes considering both its existing information systems and IT services, as well as the broader organizational context.

For this section, we used the findings from our case studies. We analysed and categorized the collected data to devise a generic architecture for the company´s service that analyses social media data and detects misinformation and polarization.

This architecture is based on social media big data in any form (voice, text, image, video) and is analysed by a specific component. This component considers industry standards, company specific regulations and expert´s insight in knowledge management. It also incorporates knowledge from the company´s databases and knowledge bases through the existing information system and technology (See Fig. 4).

Companies utilize diverse data sources to identify misinformation and polarization, drawing from operational activities, internal records, and external references such as internet sources, audio recordings, and image files. This systematic data collection and analysis enables companies to derive insights and inform decision-making processes. Initially, unstructured data from various origins undergoes organization and cleansing, leading to the creation of a reliable knowledge base repository. This repository along with the factual database repository enables the dedicated component to predict and classify comments as misinformation or polarization. By integrating various types of data, a flexible data system is constructed that allows for ongoing refinement and utilization.

Fig. 4 illustrates the types of data utilized by companies for misinformation and polarization detection and application. Dashboards facilitate the visualization of the origins of misinformation and polarization, aiding in the identification of sources of polarized and inaccurate information. A repository of various AI tools is maintained for use within the component. Robust access control measures, including adherence to GDPR regulations, are employed to safeguard data privacy, confidentiality, and availability.

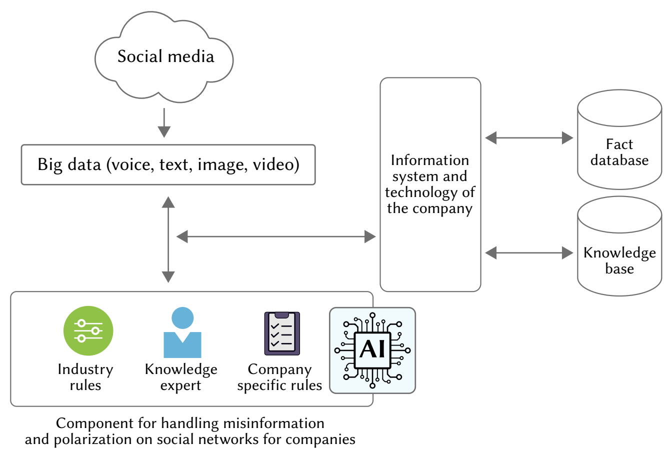


Fig. 4. Social networks misinformation and polarization component integration with existing technology for companies (Source: own elaboration).

Regarding the use of AI tools, the integration of ML techniques enhances operational resilience by improving the detection of misinformation and disinformation. Reference [37] presents a supervised ML model aimed at reducing operational interruptions, with a focus on supply chain processes. The model is praised for its role in strengthening decision-making through the assessment of diverse data sets. The study underlines the significant role of AI, and more specifically ML, in increasing the accuracy and reliability of false information identification, which is crucial for decision-making across various business specially in maintaining supply chains integrity.

However, integrating advanced AI-ML into existing systems presents significant challenges, requiring vast computational power and compatibility with established data systems [38]. The rapid evolution of AI technologies necessitates flexible and adaptable integration methods capable of managing intricate data analyses and accommodating ongoing updates. Achieving seamless integration involves not just technical alignment but also considerable investments in time, money, and specialized knowledge. Maintaining a balance between introducing innovations and preserving operational consistency is essential for organizations to fully leverage the benefits of AI-ML enhancements.

A dedicated component like this would be of great importance to tackle misinformation spreading online, and could represent a key element of a system (observatory) to constantly monitor information flow in real time, enabling the issue of warnings about topics that require special caution. Respondents R2 and R5 recommend taking necessary measures to protect customer data according to existing customer protection laws. Therefore, this component should only utilize the latest research for detecting misinformation and polarization on social media, but also operate in compliance with General Data Protection Regulation (GDPR). Moreover, organizations must take appropriate technical and organizational measures to ensure the security of personal data. This includes implementing access controls, encryption, and other security protocols to prevent unauthorized access, use, or disclosure.

Respondents R2 and R5 commented that the current enterprise systems commonly deployed in companies on average are of flexible nature, often based on use of big data, and allow existing infrastructures to evolve and adapt new components. Currently , Big data is used for other business dimensions, such as customer profiling, cost of sales, market testing, price optimization, and understanding customer preferences. The use of social media data for detecting misinformation and polarization is less frequent, but according to R1, and R7, interest in this application is growing.

One challenge identified by the responders for the implementation of architecture as proposed above for companies, consists of assessing the level of complexity of integration, and companies need to be careful when choosing components that allow easy integration. Integration with existing architecture may be difficult, expensive and time consuming and the number of hours needed to have such integration in place need to be understood prior to enrolment in such project.

R1 said "Similar components have been selected in the past based on their functional requirements and ease of integration with the already existing solution".

Another challenge for businesses lies in assessing the risk of adopting a new component. Indeed, respondents acknowledged this, and the most important factors identified were cost for adoption, ease of customization, available support, licensees, as well long-term maintenance and reassessment if the component works fine. This is in line with previous research that supports such considerations for the adoption of new tech components for companies [39] [40].

R5 described the evaluation of risk like this "when the requirements are known, and the correct component identified, a risk analysis is undertaken to understand the long-term consequences in terms of support and maintenance of the component, and other risks, such as, commercial or cost related".

R6 also stressed the need for good documentation regarding the component: "Different components exist out there, however it is important to have a component that is well-documented and easy to maintain".

Some of the responders (R4 and R5) commented that when implementing such components, there must be a balance between speed to deliver such components, and maintaining the quality and stability of the overall architecture. R5 said "Solutions in which one supports deployment only of this particular component rather than deployment of the entire system must be put in place for scenarios as this one".

Respondents R3 and R6 said that they see it important that AI models are trained with company domain-specific knowledge, and not only as in previous literature, on misinformation and polarization social media knowledge. For the moment, in previous literature [41] [42], approaches to improve misinformation and polarization rely on very basic deep learning models, not on company and domain specific knowledge, and there is a growing interest to expand knowledge on misinformation/polarization possibilities and understand how companies can run this kind of analysis on more complex deep learning models. At the same time, in the view of R3 and R6, AI models must be explainable; that is, in addition to detecting the sources of polarized and inaccurate information, they must explain to business users, how they have come up with the decisions taken, as well as why these comments and opinions are classified as misinformation and polarization compared to the others, to implement AI responsibly and transparently.

IV. Discussion

Implications from research like ours are multiple. One relevant aspect, in discussions such as we start in this work, is the ethical and social responsibility of social platforms themselves [43] [44]. The social platforms are designed, so that more users use them, the more money they make. Social media companies have vested interest to keep users as much as possible on their platforms. They have a big number of personnel hired specifically to study what captures users' attention. As also development and deployment or changes in the current outlook are easy to perform, they launch a new feature of the social media platform in matter of weeks, so these companies have reached the point in which social media become irresistible for the users. We conclude with previous research, that social media initially were not created intentionally to spread misinformation and polarization, and that exactly these engagement mechanisms and vested interests of social companies, have contributed to this growing trend [45]

Perverseness of these technologies on one side, combined with the potential manipulability of these platforms as mentioned above, suggests that corporation's ethical and social responsibilities need to be revisited and include challenges beyond what was traditionally answered with corporate social responsibility agenda of companies. This is important to recognize, and, in this work, we aim to understand better this phenomenon, that is at least to understand the factors that led to it, and what companies can do in this respect [44].

Previous research has shown how a company approaches its corporate digital responsibilities varies from organization to organization, as well as the domain in which it operates [46] [47]. These norms and values that an organization follows are influenced by public opinion legal requirements, technological progress, industry factors, customer factors, and firm factors [46]. Previous research has also suggested that organizations need to follow ethical responsibility norms, at each step of data collection and use and make impact assessments [48].

We foresee positive correlation with companies' financial performance if companies implement responsible social media presence, our work in an initial first work in the domain, and we provide initial guidance to practitioners to drivers to misinformation and polarization as well as technological factors for social media misinformation and polarization mitigation on company level. We believe that future works should provide conceptual and analytical models that assist companies and managerial decision making, involving different stakeholders within the company; continued work in this field is timely and urgently needed.

Our findings lead us to implications and conclusions from both dimensions, social and technological. From the social dimension experiment, in view of the results obtained in the experiment carried out on 215 participants to analyse the levels of polarization and misinformation, it is observed that 33.5% of the participants changed the polarization of the text they received (approximately, 1 in 3 people changed the polarization of the message they received), while 27% of the participants rewrote their texts using half or less of the keywords in the original text (approximately 1 out of 4 people was unable to retain or maintain in their rewritten text more than half of the key words of the original text that they received). This shows that polarization is a prevalent phenomenon, as a significant percentage of participants changed the polarization of the text they received. This highlights the influence of external factors on individuals' perspectives and the potential for polarization to occur. Regarding misinformation, this demonstrates that accurately retaining and conveying information is challenging and suggests a difficulty in understanding and conveying the intended message accurately.

Gender does not seem to be a determining factor in polarization, since, during the experiment carried out, women and men changed the polarization of the text they received at very similar levels (34.71% of women compared to 34.62% of men). However, a significant difference is identified between women and men in terms of misinformation. Specifically, women generated 29.75% compared to 23.4% of men, so gender does seem to be a factor to consider in the misinformation. Therefore, gender may play a role in misinformation, as there was a notable difference between women and men in terms of generating misinformation. Further exploration is needed to understand the underlying factors contributing to this disparity.

Considering age groups, a correlation is identified regarding misinformation across different age ranges, except for the 41-50 age group where it slightly decreases. Therefore, in general, an increase in misinformation levels is observed as age increases in the following ranges: 18-30 (24.22%), 31-40 (25%), 51-60 (31%), and +60 (43.48%). This could be related to the cognitive degeneration that occurs as we get older. However, in relation to polarization considering age, no correlations are observed, although there are very high levels of polarization in the age groups 41-50 (61.54%), followed by people aged +60 (47 .83%).

It can be concluded, therefore, that both age and gender are determining factors in misinformation, whereas polarization only occurs partially. Knowing this fact, work could be done in the area of misinformation to improve levels, trying to act on gender and age factors. Strategies should focus on improving information retention and promoting critical thinking skills, particularly among older individuals.

From the technical dimension experiment, we noted that the volume of information (and misinformation) regarding companies that is available online is big, and the extent to which it's used possess challenges for business from misinformation and polarization points of view for the companies is significant. The businesses whose staff participated in this study have technologies and say that suited machine learning models can be easily integrated in existing systems to analyse data and understand perception of the company, as we propose in this work. However, it is noted that to have successful implementation, there must be a human oversight, and constant monitoring and integration of current customer protection rights acts, for responsible use of social media data. Further, there is an evolving understanding and attention regarding the importance of the issue, both within the society as whole and business, that even further accelerate the developments towards this topic.

What our interviewers told us is that the process of understanding long-term maintenance and viability of such projects within companies can be hindered due to conflicting agendas and priorities in the decision making of the company, yet change is inevitable, and it is to be expected that companies when deciding on such components will look for solutions less likely to change in future. Further considerations for pace and scale of development of such solutions will need to be decided too.

Our research shows that companies' decision makers and suited personnel already have suited technical infrastructures, that easily can allow incorporation of such responsible misinformation and polarization detection components. We presented the requirements and design for a suited technology solution, and future work concerns the type, identification, assessment, and evaluation of suited ML models to use for this aim. The introduction of detection for use of misinformation and polarization has limitations in that i) the area of machine learning is only in its nascence, therefore detected results may be partially correct or sometimes maybe incorrect [41], and ii) the ethical implications for policies a company should adopt and how to behave in situation if misinformation or polarization is present and handled are not easy and straightforward task, requiring consolidation view by different stakeholders (not only the company) to develop a common understanding on how such behaviours are accounted. An even further challenge is such detection and decision making for such situations should be fully or partially automated.

As in any study, there are limitations. In this case, the social experiment was carried out starting from the university community and considering a small number of very general factors, such as age, sex, language, etc. Subsequent research where experiments will be carried out on a larger scale, or perhaps more focused on a set of specific factors to more thoroughly analyse their impact in terms of polarization and misinformation, could be of great interest to advance in this area. Finally, the text used in the present study was intended to be as neutral as possible, but in subsequent research it would be possible to work on different texts and levels of polarization. The generalizability of the findings in the technological experiment is limited in a number of ways as well. Our case studies were conducted with technological companies, that is providers of enterprise systems, which are the forerunners in implementing technology solutions for companies, and therefore are the first stakeholders willing to implement such solutions, yet such requests for solutions are expected to be on rise, due to hindering nature of the underlying problem.

Second limitation in the technological experiment comes from the qualitative nature of the study, with three technology companies interviewed, therefore future research will need to be performed on a larger scale. By highlighting the problem of presence of misinformation and polarization on social networks, especially in the reality of companies, this work shows on one side factors that lead to misinformation and polarization on social networks, and on another hand, what and how can companies approach this problem with specific technology tools developed for this aim. Future research considering different deployment models and training models can enhance the currently proposed solution. Our study of both, factors and technical solutions for misinformation and polarization for companies are therefore first and an important step, but more research is still needed.

V. Conclusions

Our work is a first study trying to understand social and technical factors with respect to misinformation and polarization for companies. We understood that one third of participants in our experiments polarizes texts on social media when needing to retell them, while older people (age 41 and above) spread misinformation more often compared to the other age groups and tend to polarize information more easily. Technological analysis says that recent research has presented relevant components for the detection of misinformation and polarization and such components are easy to develop and integrate with existing enterprise architectures of companies, with factors like cost for adopting, ease of customization, support, licenses, as well long-term maintenance being important for the adoption.

Our work can help practitioners in companies to improve their understanding regarding misinformation and polarization, provides them information on what they can practically do to implement such services, and provides awareness for important aspects for setting more sophisticated services in future.

Up to our knowledge, our work is a first attempt to understand the effects of polarization and misinformation on social media for companies and what can companies do to cope with such presence. This is an emerging research area that requires further investigation.

This study opens several avenues for future research. In light of the constantly changing dynamics of misinformation and polarization, especially within social media environments, it is crucial to delve into the incorporation of sophisticated AI and ML strategies. These strategies should be capable of adjusting to the intricate subtleties of human interaction. Future studies could focus on developing more sophisticated models that account for the dynamic context of misinformation spread and the varying degrees of polarization. Additionally, investigating the long-term impact of misinformation and polarization mitigation strategies on organizational trust and consumer behaviour could provide valuable insights for both academia and industry.

As this field continues to grow, interdisciplinary approaches combining insights from social sciences, computer science, and organizational behaviour could enhance our understanding of misinformation's multifaceted impact. Moreover, with the increasing importance of ethical considerations in AI development and deployment, future research should also focus on the ethical frameworks that guide the use of AI in combating misinformation, ensuring that these technologies respect privacy, consent, and fairness.

References

[1] M.A. Al-Asadi, A. Mustafa, S. Tasdemir, "Using artificial intelligence against the phenomenon of fake news: a systematic literature review," in *Combating Fake News with Computational Intelligence Techniques. Studies in Computational Intelligence*, vol 1001, M. Lahby, AS.K. Pathan, Y. Maleh, W.M.S. Yafooz, Eds. Springer, Cham. https://doi.org/10.1007/978-3-030-90087-8_2 2022, pp. 39-54.

[2] H. Allcott, M. Gentzkow, C. YU, "Trends in the diffusion of misinformation on social media," *Research & Politics*, vol. 6, no 2, p. 2053168019848554, 2019.

[3] V.L. Rubin, "Deception detection and rumor debunking for social media," in *The SAGE handbook of social media research methods*, L. Sloan, A. Quan-Haase, Eds. London: Sage, 2017. p. 342.

[4] V.L. Rubin, N. Conroy, Y. Chen, S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection,* T. Fornaciari, E. Fitzpatrick, J. Bachenko, Eds. 2016, pp. 7-17.

[5] J. Boase, B. Wellman, "The internet and email aid users in maintaining their social networks and provide pathways to help when people face big decisions," *PEW internet and American life project*, 2006.

[6] C. Kadushin, *Understanding social networks: Theories, concepts, and findings*, Oxford university press, 2012.

[7] E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 519-528.

[8] P.W. Kraft, N. R. Davis, T. Davis, A. Heideman, J.T. Neumeyer, S.Y. Park, "Reliable Sources? Correcting Misinformation in Polarized Media Environments," *American Politics Research*, vol. 50, no 1, pp. 17-29, 2022.

[9] P.N. Petratos, "Misinformation, disinformation, and fake news: Cyber risks to business," *Business Horizons*, vol. 64, no 6, pp. 763-774, 2021.

[10] D. Spohr, "Fake news and ideological polarization: Filter bubbles and selective exposure on social media," *Business information review*, vol. 34, no. 3, pp. 150-160, 2017.

[11] A.M. Guess, B.A. Lyons, "Misinformation, disinformation, and online propaganda," *Social media and democracy: The state of the field, prospects for reform*, N. Persily and J. A. Tucker, Eds. Cambridge: Cambridge University Press, 2020, pp. 10-33.

[12] J. Shin, L. Jian, K. Driscoll, F. Bar, "The diffusion of misinformation on social media: Temporal pattern, message, and source," *Computers in Human Behavior*, vol. 83, pp. 278-287, 2018.

[13] E.K. Vraga, L. Bode, "Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation," *Political Communication*, vol. 37, no. 1, pp. 136-144, 2020.

[14] Y. Wang, M. McKee, A. Torbica, D. Stuckler, "Systematic literature review on the spread of health-related misinformation on social media," *Social science & medicine*, vol. 240, p. 112552, 2019.

[15] P.M. Fernbach. L. Van Boven, "False polarization: Cognitive mechanisms and potential solutions," *Current opinion in Psychology*, vol. 43, pp. 1-6, 2022.

[16] A. Bessi, F. Zollo, M. Del Vicario, A. Scala, G. Caldarelli, "Trend of narratives in the age of misinformation," *PloS one*, vol. 10, no. 8, p. e0134641, 2015

[17] M.H. Ribeiro, P.H. Calais, V.A.F. Almeida, "Everything I disagree with is# FakeNews: Correlating political polarization and spread of misinformation," *arXiv preprint arXiv:1706.05924*, 2017.

[18] Y.K. Chang, I. Literat, C. Price, J.I. Eisman, J. Gardner, A. Chapman, A. Truss, "News literacy education in a polarized political climate: How games can teach youth to spot misinformation," *Harvard Kennedy School Misinformation Review*, 2020.

[19] R. Cooper, M. Foster, "Sociotechnical systems," *American Psychologist*, vol. 26, no 5, pp. 467-474, 1971

[20] G.H. Walker, N.A. Stanton, P.M. Salmon, D.P. Jenkins, "A review of sociotechnical systems theory: a classic concept for new command and control paradigms," *Theoretical issues in ergonomics science*, vol. 9, no. 6, pp. 479-499, 2008.

[21] D. Tilson, K. Lyytinen, C. Sørensen, "Research commentary—Digital infrastructures: The missing IS research agenda", *Information systems research*, vol. 21, no. 4, pp. 748-759, 2010.

[22] A. Tiwana, B. Konsynski, A.A. Bush, "Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics," *Information systems research*, vol. 21, no. 4, pp. 675-687, 2010

[23] A. MacCormack, "Product-development practices that work: How Internet companies build software," *MIT Sloan Management Review*, vol. 42, no. 2, p. 75, 2001.

[24] P. Rodríguez, A. Haghighatkhah, L.E. Lwakatare, S. Teppola, T. Suomalainen, et. al. "Continuous deployment of software intensive products and services: A systematic mapping study," *Journal of systems and software*, vol. 123, pp. 263-291, 2017.

[25] H.H. Olsson, J.Bosch, H. Alahyari, "Towards R&D as innovation experiment systems: A framework for moving beyond agile software development," in *Proceedings of the IASTED*, pp. 798-805, 2013.

[26] S. Bellomo, P. Kruchten, R.L. Nord, I. Ozkaya, "How to agilely architect an agile architecture," *Cutter IT Journal*, vol. 27, no. 2, pp. 12-17, 2014.

[27] N. Brown, R.L. Nord, I. Ozkaya, M. Pais, "Analysis and management of architectural dependencies in iterative release planning," in *2011 Ninth Working IEEE/IFIP Conference on Software Architecture.* IEEE, 2011. pp. 103-112.

[28] S. Monteith, T. Glenn, J.R. Geddes, P.C. Whybrow, E. Achtyes, M. Bauer, "Artificial intelligence and increasing misinformation," *The British Journal of Psychiatry*, vol. 224, no. 2, pp. 33-35, 2024.

[29] S. Kreps, R.M. McCain, M. Brundage, "All the news that's fit to fabricate: AI-generated text as a tool of media misinformation." *Journal of experimental political science, vol.* 9, no. 1, pp. 104-117, 2022.

[30] Q. Su, M. Wan, X. Liu, C.R. Huang, "Motivations, methods and metrics of misinformation detection: an NLP perspective," *Natural Language Processing Research*, vol. 1, no. 1-2, pp. 1-13, 2020.

[31] Meta. "Here's how we're using AI to help detect misinformation". Accessed: Feb. 15, 2024. [Online]. Available: https://ai.meta.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/

[32] E. Breck, C. Cardie, "Playing the telephone game: Determining the hierarchical structure of perspective and speech expressions," in *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, 2004, pp. 120-126.

[33] F. Breithaupt, B. Li, "Fact vs. affect in the telephone game: All levels of surprise are retold with high accuracy, even independently of facts," *Frontiers in psychology*, vol. 9, p. 375712, 2018.

[34] E.J. Mulder, D. Snijders, "Playing the telephone game in a multilevel polity: On the implementation of e-government services for business in the EU," *Government Information Quarterly*, vol. 39, no. 2, p. 101526, 2022.

[35] G. Walsham, "Interpretive case studies in IS research: nature and method," *European Journal of information systems*, vol. 4, no. 2, pp. 74-81, 1995.

[36] R.K. Yin, "Designing case studies," *Qualitative research methods*, vol. 5, no. 14, pp. 359-386, 2003.

[37] P. Akhtar, A.M. Ghouri, H.U.R. Khan, M. Amin ul Haq, U. Awan et al. "Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions," *Annals of Operations Research*, vol. 327, no. 2, pp. 633-657, 2023.

[38] J. Tello, L. Subramaniam, "The magic behind turning data into profit," *Deloitte.* Accessed: Feb. 15, 2024. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-challenges-of-using-artificial-intelligence-final.pdf

[39] S. Butler, J. Gamalielsson, B. Lundell, C. Brax, A. Mattsson et al. "Considerations and challenges for the adoption of open source components in software-intensive businesses," *Journal of Systems and Software*, vol. 186, p. 111152, 2022.

[40] A. Davila, G. Foster, "Management accounting systems adoption decisions: Evidence and performance implications from early-stage/startup companies," *The Accounting Review*, vol. 80, no. 4, pp. 1039-1068, 2005.

[41] D. Caled, M.J. Silva, "Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation," *Journal of Computational Social Science*, vol. 5, pp. 123–159, 2022.

[42] D. Cavaliere, G. Fenza, V. Loia, F. Nota, "Emotion-Aware Monitoring of Users' Reaction With a Multi-Perspective Analysis of Long- and Short-Term Topics on Twitter," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 8, no. 4, pp. 166-175, 2023, https://doi.org/10.9781/ijimai.2023.02.003

[43] M. Sandoval, *From corporate to social media: Critical perspectives on corporate social responsibility in media and communication industries*, Routledge, 2014.

[44] C. Stohl, M. Etter, S. Banghart, D.J. Woo, "Social media policies: Implications for contemporary notions of corporate social responsibility," *Journal of business ethics*, vol. 142, pp. 413-436, 2017.

[45] S. Vosoughi, D. Roy, S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146-1151, 2018.

[46] L. Lobschat, B. Mueller, F. Eggers, L. Brandimarte, S. Diefenbach et al. "Corporate digital responsibility," *Journal of Business Research*, vol. 122, pp. 875-888, 2021.

[47] B. Mueller, "Corporate digital responsibility," *Business & Information Systems Engineering*, vol. 64, no. 5, pp. 689-700, 2022.

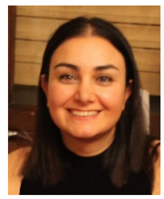
[48] L. Martín-Gómez, J. Pérez-Marcos, R. Cordero-Gutiérrez, D. H. De La Iglesia, "Promoting Social Media Dissemination of Digital Images Through CBR-Based Tag Recommendation," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 7, no. 6, pp. 45-53, 2022. https://doi.org/10.9781/ijimai.2022.09.002

Alberto Tejero

Dr. Alberto Tejero holds an Master Degree in Business Administration - MBA (Universidad Complutense de Madrid), an MSc in Computer Science (specializing in Cybersecurity), and a PhD in Telecommunications Engineering from Universidad Politécnica de Madrid (UPM). He is an Associate Professor at UPM, Director of the Master in Digital Innovation (UPM and European Institute of Innovation and Technology), former Deputy Director of the Transfer Office at UPM and an expert in international Technological Innovation Management, with a focus on entrepreneurship and innovation ecosystems. Alberto Tejero advises Foundations and Councils on innovation and education, and actively participates in international innovation and entrepreneurship projects in Europe, Latin America, and the USA.

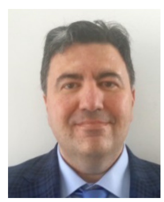
Galena Pisoni

Dr. Galena Pisoni is a Lecturer (Assistant Professor) in Data Analytics at York Business School, York St John University, and a Visiting lecturer at IAE School of management, Université Côte d'Azur, Nice, France. Her research interests include enterprise information systems, data science, and knowledge management for companies of different sizes. She holds a Ph.D. in Computer Science with a focus on technology design and human-centered design of socio-technical systems. In her past she was a visiting researcher at center for Entrepreneurship, Strategy and Innovation management, University of Twente, Faculty of Industrial Design Engineering, Delft University of Technology, School of Architecture (Smart Spaces), KTH, and the Netherlands Organisation for applied scientific research (TNO), Groningen. She has an academic track record in business informatics and business intelligence, with publications in international conferences and journals in the broad areas of information systems, business and innovation.



Ziba Habibi Lashkari

Dr. Ziba Habibi Lashkari is an Associate Professor in Finance at Universidad Politécnica de Madrid. She has elevated her academic career to this distinguished position, showcasing her expertise and dedication to the field of FinTech. She currently serves as the Coordinator of the Master of FinTech program in EIT Digital and is also the Coordinator of the EIT Digital Summer School. She holds a Ph.D. in Economics and received the Ph.D. Extraordinary Award in 2019. She also is the coordinator of the "FinTech" summer school in Madrid and has been recognized for her innovative ideas in cybersecurity for fintech, earning a diploma for best startup ideas at the 18th ActúaUPM. A member of the "Decision Analysis and Statistics Group (DASG)".



Sergio Rios Aguilar

Dr. Sergio Ríos holds a PhD in Computer Science (2021, URJC, Spain), a PhD in Economics and Business Sciences (2013, University of Granada, Spain), and a MSc in Telematics and Information Systems (URJC, 2008, Spain). He is an Associate Professor and Deputy Director of the Computer Science School at Universidad Politécnica de Madrid (UPM). He is a Lecturer in Emerging Technologies and Business Innovation, as well as Lecturer in I&E at the European Institute of Technology (EIT) MSc Programme. He was also the Director of a recognized Research Group on Mobility and User Experience (MUX) for 4 years. He has worked as a consultant in projects for several international companies like Samsung Electronics, Nokia, INTU Group, and others. He is a member of the European Global Navigation Satellite Systems Agency's (GSA) Galileo Raw Measurement Taskforce. His current research interests include AI-powered Business Information Systems, Mobile presence and attendance systems and Indoor location-based Services.