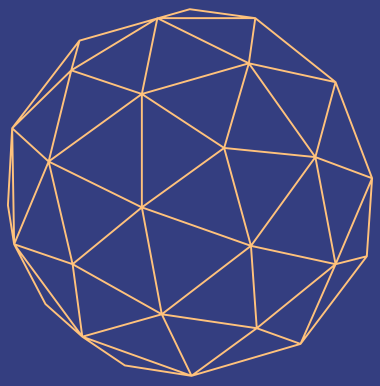
Issue 2 - November 2021

Good In Tech RESEARCH NEWS

Rethinking innovation and technology as drivers of a better world for and by humans



Artificial Intelligence social impact

AHMAD HAIDAR, CHRISTINE BALAGUÉ

Responsible Innovation in AI: A Systematic Literature Review & Guidance for Future Research

MAXIME CRÉPEL, DOMINIQUE CARDON

Textual data analysis of corpora on ethics and disorders related to AI and algorithms

NESMA HOUMANI, CHRISTINE BALAGUÉ

EXplainable Artificial Intelligence systems for Health

Institut Mines-Télécom Business School
TELECOM Paris TELECOM SudParis INSTITUT POLYTECHNIQUE DE PARIS
FONDATION DU RISQUE Louis Bachelier
SciencesPo ÉCOLE DU MANAGEMENT ET DE L'INNOVATION

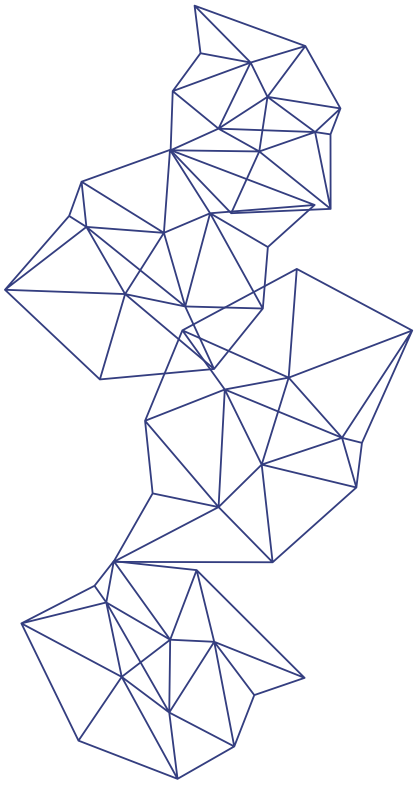
GOOD IN TECH VISION

Good In Tech main objectives are to create knowledge around four research areas and to contribute to the dissemination of this knowledge not only in academic and pedagogical spheres but also to corporations, decision-makers, regulators and the general public.

To this end, the Chair aims to create and develop an ecosystem of interactions between research, companies, students from the two partner academics and political institutions, civil society in order to raise awareness of all stakeholders on this new paradigm on responsible digital technologies and innovation.

The chair also aims to develop international partnerships, particularly in Europe, to share the issues of responsible digital innovation with international committees.

Finally, the Chair aims to share the results of academic works and debates it organizes with national and European political institutions in order to inform and influence public policies.



Responsible Innovation in AI: A Systematic Literature Review & Guidance for Future Research

Ahmad Haidar, Christine Balagué

Ahmad Haidar

Ahmad is a PhD researcher at Paris Saclay University's LITEM laboratory and sponsored by the Good In Tech Chair since November 2020. His research supervised by prof. C. Balagué focuses on sustainable impact of Artificial Intelligence, using econometric models. Ahmad holds a dual Master's degree in Economics from Lebanese University and Business Management. His memoir is titled "The Impact of Digital Transformation and Artificial Intelligence on Economy." He also assisted in developing a database and data processing for the American University of Beirut project as a research assistant. Additionally, as a trainee at the Central Bank of Lebanon and other alfa banks.

Christine Balagué

Christine Balagué is HDR Professor at IMT-BS and holder of the Good in Tech Chair (www.goodintech.org). Her research focuses on modelling the behaviour of connected individuals, ethics of technology and AI, and responsible digital innovation. She is also a member of several national committees: CSA expert committee on online disinformation, Defence ethics committee, Haute Autorité de Santé recommendations impact commission, executive committee of Cap Digital. As Vice-President of the National Digital Council from 2013 to 2016, she is also co-author of several reports submitted to the French government on digital issues. She published more than 50 research articles & international conference proceedings as well as several books on society and economic digital metamorphosis.

WHY IS THIS TOPIC IMPORTANT ?

This paper investigates **the relationship between artificial intelligence and responsible innovation in management**. AI is also considered as a management challenge, engaging leaders and executers to promote the ethical use of technology and to manage responsibility towards those affected by the changes of AI. Employees need to acquire new skills and the know-how to navigate the transitions of AI and prepare for it.

Responsible innovation framework helps to guide the governance of Artificial Intelligence, according to **Stilgoe, Owen & Macnaghten's paper (2013)** Developing a framework for responsible innovation. Their framework for responsible innovation consists in **four dimensions: anticipation, reflexivity, inclusion and responsiveness**.

Additionnally, according to **Thomas Hemphill (2020)**, there are **four innovation governance approaches** for innovaton, requiring different regulation and strategies: **precautionary principle, responsible innovation,permissionless innovation, principle of innovation**.

Companies use Artificial Intelligence to solve complex tasks using a huge amount of data and machine learning models. The objectives are principally to make people's lives and organizational tasks easier and faster, to improve productivity or to decrease costs. However, any framework businesses adopt to achieve these goals, companies also try to implement ethical AI criteria: robustness, respect for autonomy, fairness, explicability, democratic participation, and risk management. Indeed, mistakes might happen, that can be splitted into three main parts:

- **Mistakes during the preprocessing phase**, specifically data bias and potential discrimination effects
- **Mistakes during the learning phase** include intentional (or unintended) unfairness embedded in algorithms, privacy violation, and others.
- **Mistakes during the operational or performance phase**, conducting to a problematic usage of the technology with consequence son individuals or society.

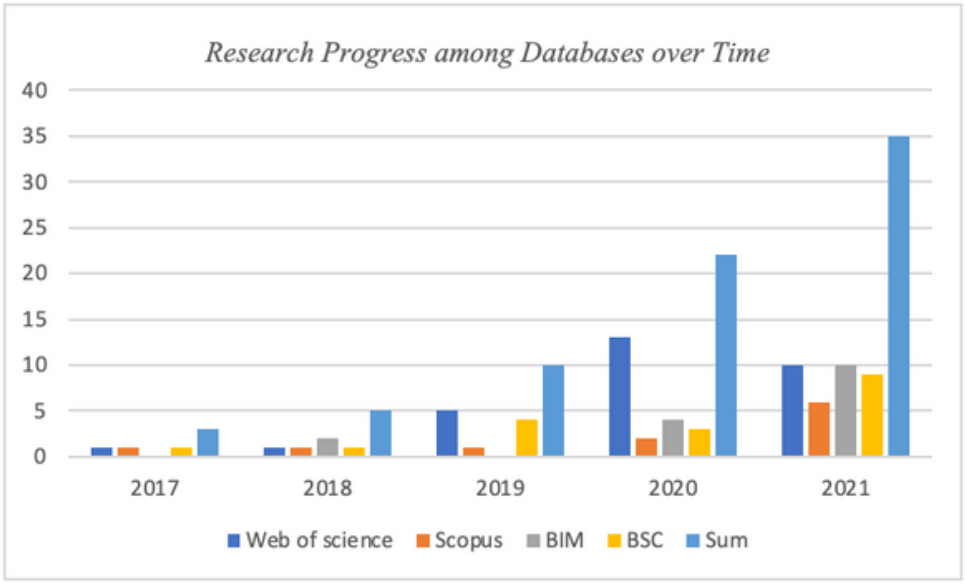
The topic of this paper is to make a Systematic literature Review on responsible innovation and artificial intelligence in management discipline, in order **to identify the most important challenges companies must face to implement AI** with a responsible innovation management framework.

METHODOLOGY

The methodology is a systematic literature review between 2017 and 2021 on artificial intelligence and responsible innovation. The authors followed Tranfield et al.'s (2003) three-stage procedure: planning, execution, and reporting.

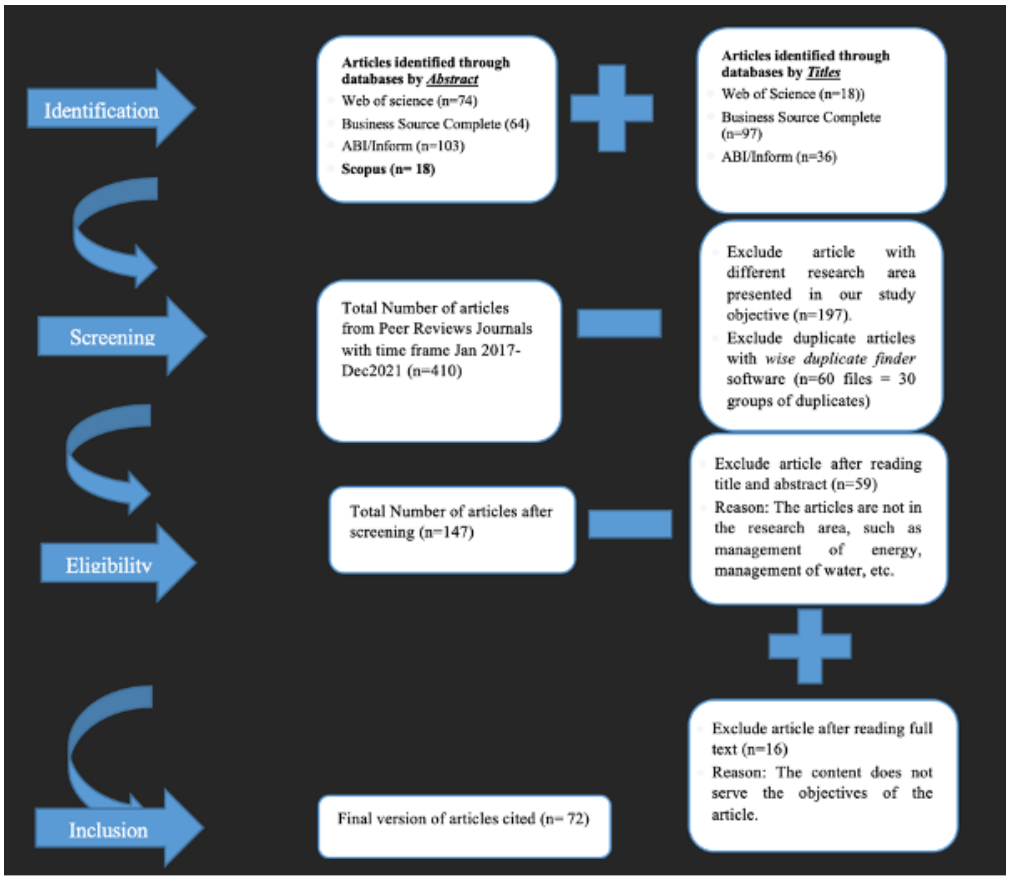
- **Planning review:** trying to formulate the problem, develop a protocol, and validate the protocol used.
- **Execution review:**
  - **Search the literature** using inclusion criteria and exclusion criteria (for example, "2017-2021", "English," "type of document: article") to select the articles needed and analyze them. This resulted in 72 articles.
  - **Assert the quality** of the articles chosen.
  - **Extract data:** the same keywords were used in the four databases: "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" AND "Responsible Innovation."
  - **Analyze the synthesis:** most of the articles found were based on a qualitative approach. So, the authors analyzed and interpreted them.
- **Reporting review:** the authors accumulate the number of articles published in 2021, higher than 2020, and higher than 2019, revealing the progression of research on AI and responsible innovation in management in the last five years.

Figure 1: Evolution Progress of Research in This Subject among Databases



Then, the authors split these 3 phases into eight main steps, with the approach used by Xiao and Watson (2019). These steps are: formulate the research problem, design and validate the review methodology, search the literature, screen for inclusion, assess quality, obtain data, analyze and synthesize data, and report the findings. The following scheme shows four main steps: identification, screening, eligibility, and inclusion.

Figure 2: Research Flowchart Design



KEY FINDINGS

The review of the analysis of results for an interval of time between 2017 and 2021, followed by discussion reveals three main pillars to implement AI technologies in management with responsible innovation: planning the AI, managing the authentic mirror of AI, and **responsiveness with a responsible approach for executives and employees.**

The collaboration among AI, decision-makers and employees must be arranged with a responsible governance framework. In Stilgoe's paper "Developing a framework for responsible innovation," published in 2013, the authors articulate and explore four integrated dimensions of responsible innovation: anticipation, reflexivity, inclusion, and responsiveness. The authors did the same thing but separated it into three main steps:

- **Anticipation:** do a plan with artificial intelligence. It's necessary to understand the dark side of artificial intelligence, the negative impact. But also to understand the importance of trust. It seems essential to evaluate human-AI cooperation, acceptance levels, resistance, and impact areas must be studied (Klumpp, 2018).
- **IA impact/reflexivity:** AI is impacting the process by sharing new kinds of algorithms. The second impact is: how to decide and who is going to decide? The question frequently arises of decision-making (and therefore responsibility) between the human and the robot in fintech. Who would be accountable when the company impact goes wrong? This is the reason why companies need AI risk management.
- **Responsiveness:** how should we act? The authors separate two main points: executors and employees.
    - **The executors:** various considerations are to be taken into consideration by identifying the risks resulting from AI, digital inclusion, and choosing governance approaches for IA responsibility. What kind of responsibility are now on executors? Many principles have been defined, such as AI fairness, but very few papers talk about the roles of leaders and employers. The role of executors and managers in implementation of responsible AI is fundamental.
    - **Employees:** AI destroyed different types of jobs, but at the same time, it has built new job careers. Employees require new skills, especially technological knowledge, which are essential for companies. So, it's important to understand the knowledge management system. Moreover, employees must understand the leadership role.

KEY TAKEAWAYS

- The relationship between AI and responsive innovation must be more explored in the management area. Not only in business or marketing, but in the whole management.
- Managers must understand that their responsibility and that responsible innovation is a challenge.
- Employees must understand there are a new training and skills required. For example, some compagnies such as use AI to hire their employees or use Virtual Reality for applications and trainings. The companies must get prepared for it.
- Any shift from any socio-technical system to another socio-technical system must be accompanied by two main actions: responsibility and sustainability.

Textual data analysis of corpora on ethics and disorders related to AI and algorithms

Dominique Cardon, Maxime Crépel

Dominique Cardon

Dominique Cardon is a Sociology professor at Sciences Po. Since 2010, his work has focused on the sociological analysis of web and big data algorithms in order to understand both the internal form of computation and the world that computers project on our societies. His research also focuses on the transformations of the media space and the new circuits of digital information. He is a member of the editorial board of the journal Réseaux and of the prospective committee of the CNIL.

Maxime Crépel

Maxime Crépel is a sociologist and research engineer at the Medialab of Sciences Po. His research focuses on the uses of digital technology and is partly financed by Good in Tech. He is part of the algoglitch project which aims to explore representations and forms of negotiation between users and algorithms.

WHY IS THIS TOPIC IMPORTANT ?

Algorithms are part of our daily lives in all technical devices and many researchers and experts discuss the ethical principles related to them. These observations remain sometimes out of touch or detached from our localized lives and the intimate and daily relationships we have with algorithms.

These expert reports remain essential to set out the main ethical principles and issues. Here, the question is more about **what ordinary users really want and accept or not in their relationship with algorithms**. Ultimately, how do we want to be calculated?

Maxime Crépel explores this question through the Algopress module within the Algoglitch project. **The idea is to identify the controversies produced by these technologies that are mediated in the press**, the technical and human agents involved and the troubles they are accused of bringing to society.

> **The press is full of cases in which criticism of algorithms or platforms is expressed.**

Why choose the press as a place to explore these issues?

The press is full of cases in which criticism of algorithms or platforms is expressed and we can discover by mirror effect what people hope their relationship with AI will look like. The choice of the press is methodological to **capture top-down expert discourse** as well as **bottom-down small stories** or cases brought by whistleblowers and ordinary users. The press is the ideal place to explore these two approaches.

An example

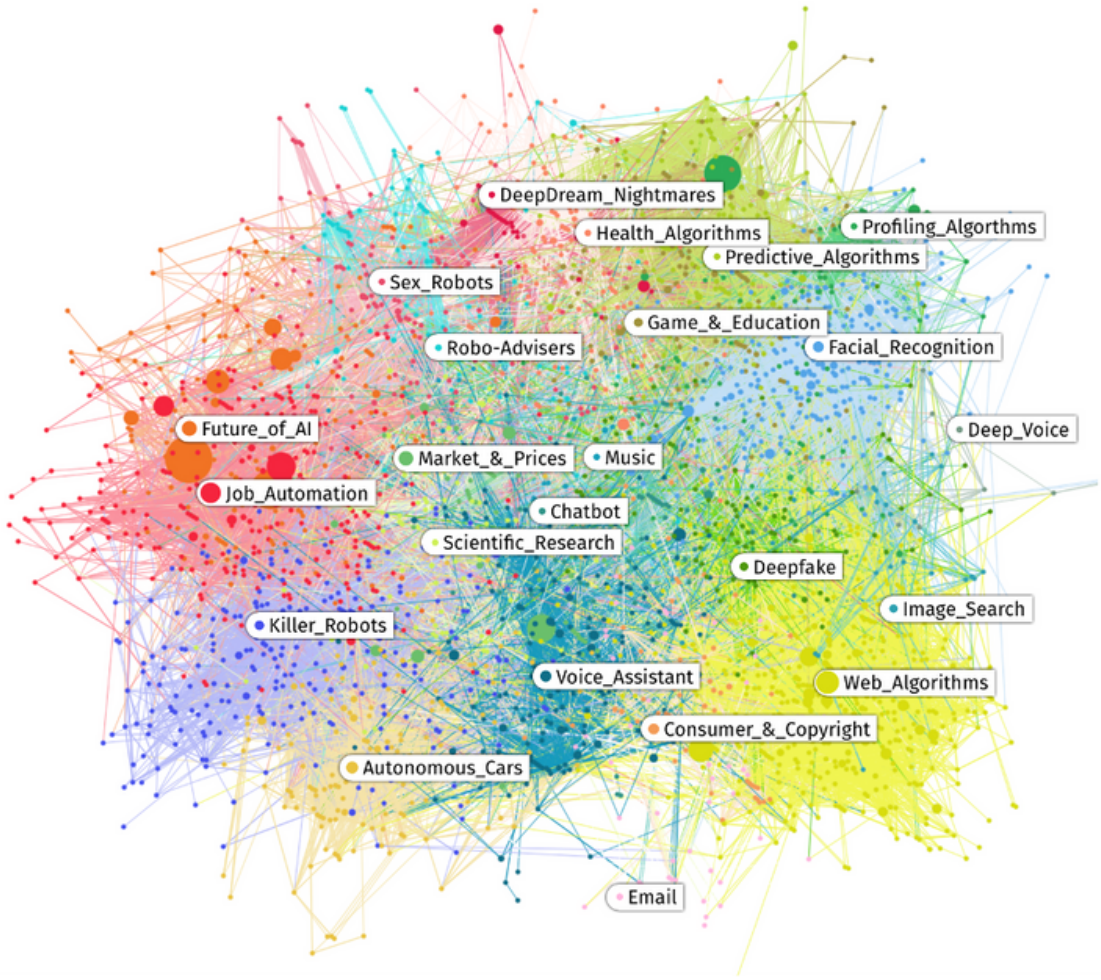
During the London Bridge attacks in 2017, Uber's algorithm, reacted to the multitude of user requests with a sharp price increase, although people were trying to leave the danger zone. In this case, the numerous complaints made a lot of noise and some journalists took the case to give it visibility in the press. Thus, the press gives us access to the speeches of ordinary users who experience difficulties in their relationship with algorithms and artificial intelligence on a daily basis.

METHODOLOGY

The authors wish to build a **semantic topology of the critical speech in the press**.

- **29,342 articles on AI and algorithms over 5 years** using Factiva press aggregator on a set of 47 general press sources in the United States and the United Kingdom (27 sources for USA and 20 sources for UK).
- Machine learning model with manual control of results
- Over **2,091 articles** detected as carrying a **critical statement about AI**
- Network of 2,991 terms criticizing AI or algorithms from which 23 thematic "clusters" emerge

Illustration : Generic map with clusters



KEY FINDINGS

Two dominant discourse regimes co-exist within media criticism.

Encart 1 : Inequalities caused by algorithms

- Critical discourse on bias, discrimination, surveillance and censorship.
- These criticisms are recognizable by the use of specific verbal groups such as "censor", "amplify", "manipulate", "discriminate »
- The victims described are particular groups of people such as facebook users, women, communities
- These inequalities are set in a present time frame

> **By reading the graph layer by layer, we can distinguish two different discourse regimes that coexist and reveal two quite different challenges**

Encart 2 : Fear of AI and robot autonomy

- Apocalyptic vision of machines overcoming human capabilities.
- In this case, the authors use the following verbs :"kill", "transform", "dominate", "replace"
- The victims described are not specific groups but all of Human kind
- These criticisms are set in the distant future

KEY TAKEAWAYS

This project answers to different goals:

- Enrich the way in which the ethical stakes of AI are usually defined by exploring the variety of actors and arguments put forward in the controversies around AI and algorithms
- Study the way in which media narratives around AI spread and feed the depictions of these technologies among the general public
- Provide data visualization devices to explore and discuss collectively the forms of acceptability and our relationship with these technologies in order to empower users.

EXplainable Artificial Intelligence systems for Health

Nesma Houmani, Christine Balagué

Nesma Houmani

Nesma Houmani is an Associate Professor at the Electronics and Physics Department of Telecom SudParis (TSP). Her skills cover a large spectrum of Artificial Intelligence domain, ranging from electronics to data science, along with signal and image processing, machine learning, pattern recognition and computer science.

Christine Balagué

Christine Balagué is HDR Professor at IMT-BS and holder of the Good in Tech Chair (www.goodintech.org). Her research focuses on modelling the behaviour of connected individuals, ethics of technology and AI, and responsible digital innovation. She is also a member of several national committees: CSA expert committee on online disinformation, Defence ethics committee, Haute Autorité de Santé recommendations impact commission, executive committee of Cap Digital. As Vice-President of the National Digital Council from 2013 to 2016, she is also co-author of several reports submitted to the French government on digital issues. She published more than 50 research articles & international conference proceedings as well as several books on society and economic digital metamorphosis.

WHY IS THIS TOPIC IMPORTANT ?

Nesma and Christine are interested in the field of **e-health and ethics**, a field that encompasses many issues and challenges that people coming from the techno or engineering world are qualified to answer.

Why is the study of ethics in algorithms very important?

Engineers working in the field of artificial intelligence develop algorithms that they understand mathematically. These can be biased when applied to different databases, when they are not operated in the same way, or when users do not have the expertise to use them properly. This can be dangerous, especially in the medical field. It is therefore essential for any user to understand the algorithm, how it works and its shortcomings, and to make sure that it is well adapted to the data on which it will be applied and to the intended objectives.

> **It is essential for any user to understand the algorithm, how it works and its shortcomings.**

Nesma is interested in the **analysis of brain activity through electroencephalography or EEG**. The EEG captures electrical brain signals. Doctors ask patients to enter a certain cognitive state in order to measure their brain activity, but there is no guarantee that the person is actually in that state. So we do not know the quality of the captured data and this adds to the complexity.

Moreover, it is very difficult to visually analyze an EEG signal because there are no patterns that stand out easily to the naked eye for a health professional, especially in case of specific neurodegenerative pathologies as Alzheimer's disease. The analysis is very time consuming.

It is therefore necessary to **develop automatic decision support tools and methods for the diagnosis of certain neurodegenerative pathologies**. Nesma's research work falls within this framework.

Nesma exploits the signals captured by EEGs and is in charge of developing algorithms that extract and analyze the markers of pathologies. She works with doctors to help them in their diagnosis and possibly determine the severity of the pathology through automatic signal processing and machine learning methods.

Within the Chair, the authors are interested in interpretation and trying to understand if the algorithms are measuring what we want and what we expect in order to **avoid biases**. Markers and work methods have already been developed, and it is now a question of understanding why these markers work and how to improve the discrimination rate of the markers. **Nesma is taking care of the technical part and Christine is trying to answer the ethical challenges.**

METHODOLOGY

The patients analyzed are geriatric ward patients and the EEG is part of their examinations and with their consent, it is exploited for research.

The researchers **apply Markovian models to model the EEG signals**. The information is quantified over epochs of the signal and they determined an average from the different epochs. The average is derived from a breakdown by time epochs. We must therefore try to understand the signal at the epoch scale.

A marker is a numerical value associated with the EEG of a person, it is a descriptor of the signal based on the Markovian model. It can be considered as a measure of synchronization between different signals.

An example with Alzheimer's disease

Nesma has developed **a method to obtain a numerical value on the information content of the EEG signal of several individuals**. Between healthy people and those with Alzheimer's disease, this marker is not activated in the same way. In order to do so, they used an average value.

The work of **analyzing and interpreting** the marker within the chair:

- Does this average value make sense? Is it representative? What does it really do?
- If we wanted to refine the diagnosis and work locally on signal epochs, how would this information content be distributed over these epochs?
  - The idea would be, for example, to work on 8ms epochs.
  - Would this effectively refine the analysis compared to using the average over the entire signal?

The authors also conduct **interviews with clinicians** in order to offer them interfaces that truly meet their needs and expectations. They listen to them to determine what they need to work with. Their responses are used to guide the research.

KEY FINDINGS

There are **different biases**:

- General, machine learning related
- Tools, interpretability and comparability of markers. Markers may have limitations that are important to identify.
- Bias regarding pre-specified diagnoses. Instead of starting from the labels that have been highlighted, it is important to question the diagnoses to create accurate markers.

From the interviews:

- Expectations and needs are very different from one practitioner to another.
- It depends on what they want to do with the tool and the pathologies they are looking for
- Many mention the need for a visualization tool as a decision aid rather than an automatic decision-making tool

KEY TAKEAWAYS

- It is therefore important, especially since the health sector is sensitive, to **set up intelligible and explainable systems** so that they can be adopted in a reassuring and reliable way by doctors who are the end-users of these technologies.
- All markers must be used with caution and technically, the methodology is to observe every detail of the signal very finely.
- This project focused on ethics and explainability allowed Nesma to address issues that did not seem urgent, in particular what the Markovian model does and the importance of the finest possible analysis. The fine-grained analysis requires a different methodology and yields very different results than those obtained using the mean.
- The interviews with practitioners revealed their particular needs, which may be the focus of further research.

Issue 2 - November 2021

Good In Tech RESEARCH NEWS

Rethinking innovation and technology as drivers of a better world for and by humans

Christine Balagué
Professor at Institut Mines-Télécom Business School
Co-holder of the Good In Tech Chair
christine.balague@imt-bs.eu

Dominique Cardon
Professor at Sciences Po and director of the medialab
Co-holder of the Good In Tech Chair
dominique.cardon@sciencespo.fr

Jean-Marie John-Mathews
Data scientist
Coordinator of the Good In Tech Chair
jean-marie.john-mathews@imt-bs.eu

Jade Vergnes
Writer for Good In Tech Research News
jade.vergnes@sciencespo.fr

Clic here to contact