

The need of trust in predictive artificial intelligence

Mapping Controversies 2022

Jorge Luis Braojos Morillas

Matilde Ficozzi

Rasmus Mølgaard Hansen

Tobias Pedersen

THE NEED OF TRUST IN PREDICTIVE ARTIFICIAL INTELLIGENCE

Exploring the controversy in research practices between predictive monitoring and trustworthy AI

What is predictive monitoring?

Predictive monitoring is concerned with anticipating the future behavior of running process instances. These instances are collected through process mining, which records events that leave traces to later on be observed and described in terms of labels.

What is Trustworthy AI?

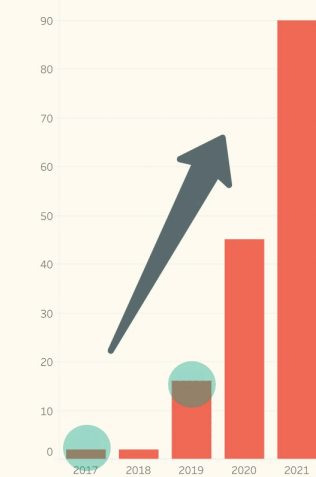
Based on the European Commission's report titled 'Ethics guidelines for trustworthy AI', the concept of trustworthy AI provides a clear benchmark for evaluating responsible development of AI solutions.

The Controversy

To what extent are researchers in predictive monitoring technologies interested in how humans are affected, is a question we have asked ourselves. AI solutions are known to exhibit inherent biases and view people as data points.

TRUSTWORTHY AI AND PREDICTIVE MONITORING ARE EMERGING TECHNOLOGIES

Trustworthy AI timeline



Trustworthy AI and predictive monitoring are both emerging fields, seeing a significant rise in published articles in recent years.

The first published literature on trustworthy AI in 2017 raises ethical questions on AI, referring to sets of practices and principles in engineering literature that may prevent the development of AI systems that could be misused. The amount of literature increases significantly in 2019 and afterwards when the European Commission published their report 'Ethics guidelines for trustworthy AI', which presented ethical requirements that AI systems should meet in order to be deemed trustworthy. The majority of research in trustworthy AI is mainly based in the subject areas computer science, mathematics and engineering. Literature from the field of social sciences becomes more prominent after 2019 suggesting that the European Commission's ethical guidelines sparked discussion and visibility on the significance of what makes AI trustworthy.

Predictive monitoring timeline

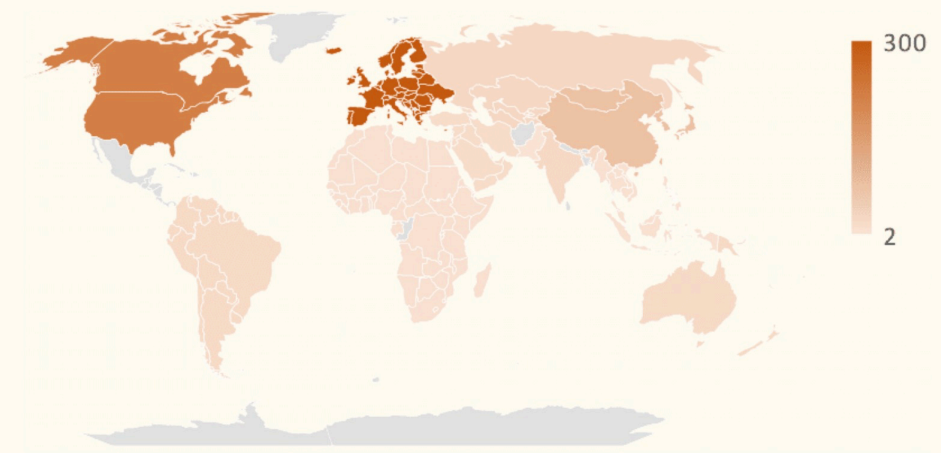


Predictive monitoring was first mentioned in 1977 in the context of identifying the electric demands of a newsprint mill, but the starting point of research in the technology emerged steadily during the period of 1985 and onwards, with a significant increase in 2010.

Early literature on predictive monitoring ranging from 1985 to 1991 shows that the concept originates in a variety of fields, mainly engineering and medicine, where literature in the field of medicine is mainly concerned with preventative care, and engineering literature widely consists of the identification of maintenance needs, systems diagnosis, and efficient manufacturing. This shows a tendency that predictive monitoring emerged in different fields for a variety of different solution purposes.

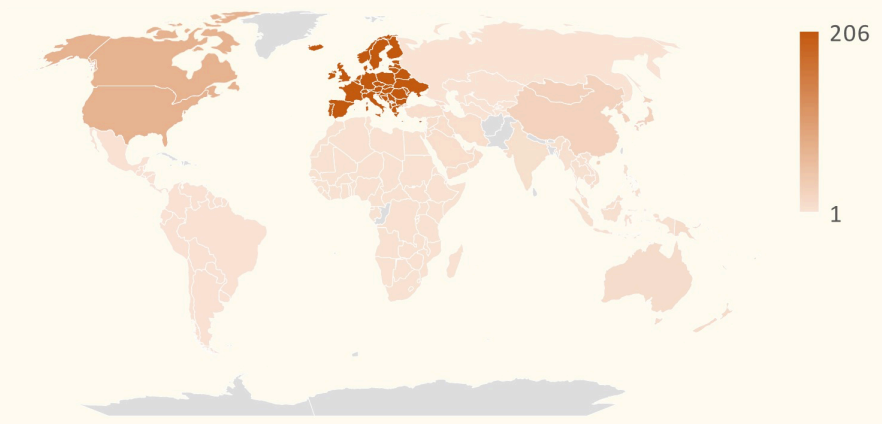
In 2010 there is an increase of literature emerging from the subject area of computer science, which is steadily interacting with articles from multiple research fields such as medicine, business, and engineering. Starting in 2011, articles are published that analyse and compare existing predictive methods across different subject areas, suggesting an interest in using predictive monitoring widely across certain research fields. We can surmise that the growing research interest in predictive monitoring coincides with how technological advances make available an increasing amount of data in a broad societal context. Early literature on predictive monitoring was limited to niche maintenance and engineering needs, but has in recent years rapidly moved into a wide variety of research areas.

THE LITERATURE IS CENTRED IN EUROPE AND NORTH AMERICA



Predictive monitoring world map

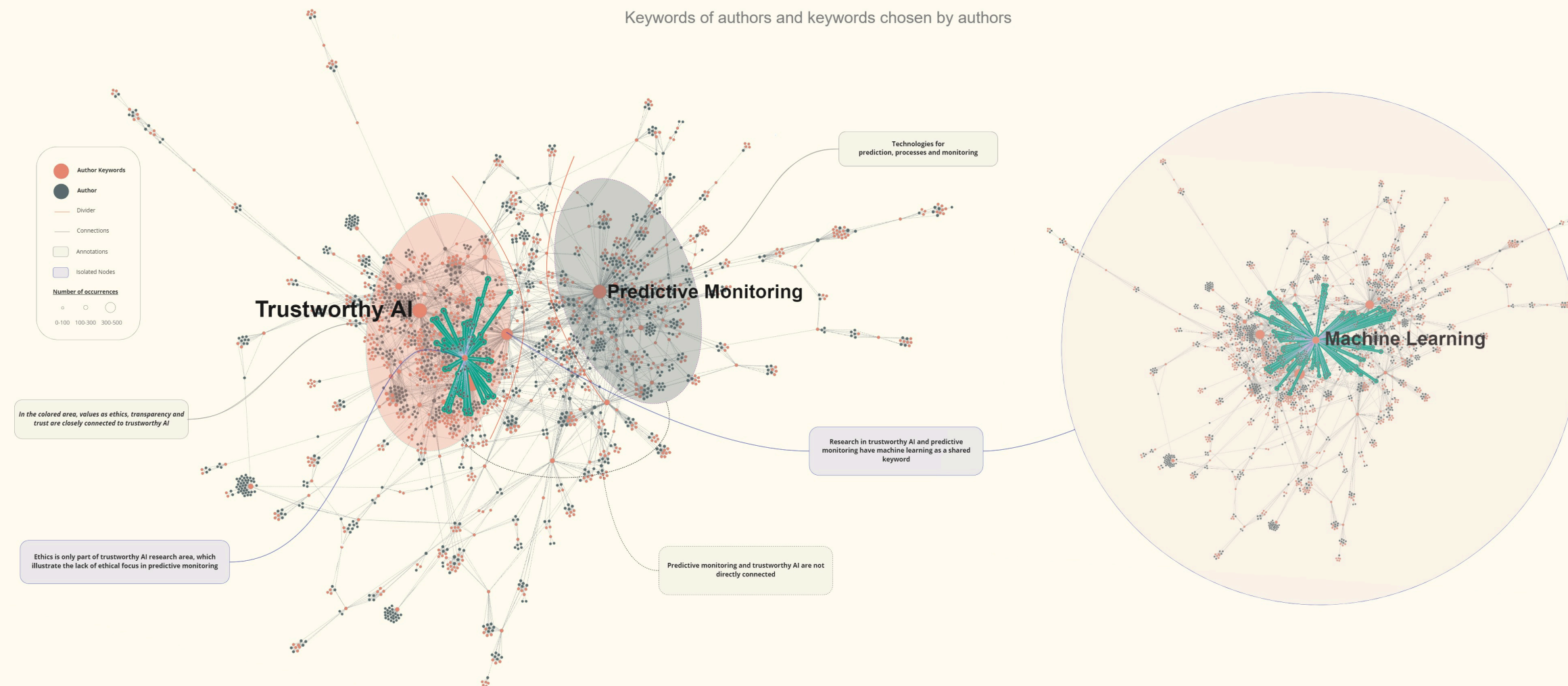
As part of our analysis placing where predictive monitoring and trustworthy are being researched in the academia, we have grouped countries in regions in the world where most research is being done, which is presented worldwide, primarily concentrated in western countries across Europe and North America for both fields.



Trustworthy AI world map

IS PREDICTIVE MONITORING WITHOUT ETHICS?

Keywords of authors and keywords chosen by authors



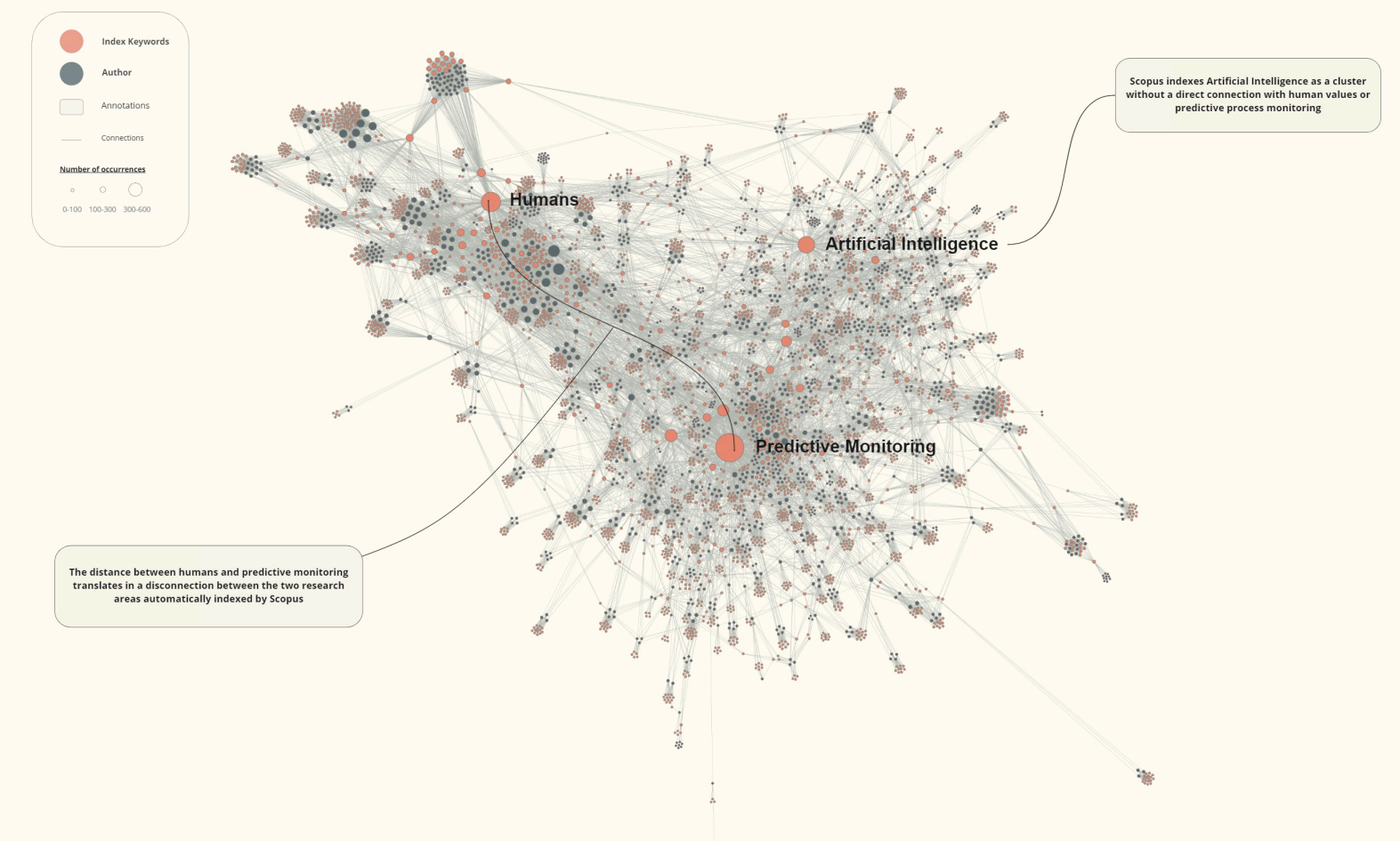
This visualisation shows the relation between keywords used in research on the topics of AI and predictive monitoring by extracting data from the scientific database Scopus. The keywords were connected with the authors in a so-called bipartite network where more frequently used keywords appear proportionately larger. The main finding shows how trustworthy AI and predictive monitoring are not interconnected directly. This gap between them illustrates how they do not mention one another, but more importantly, how the human research area is not sufficiently discussed in the research on predictive monitoring. When zooming into the network, we see how keywords such as ethics are only referred in the research about trustworthy AI, which is problematic because of predictive monitoring's impact on citizens.

The lack of interest in understanding how the human is going to be affected by predictive monitoring is alarming. Both because this technology is receiving more attention, and the fact that several components of these processes and prediction models monitor how humans interact with technology, which translates into a direct impact on their lives.

For instance, from a more technical point of view, an interesting finding through this visualisation was observing how machine learning was overlapping between trustworthy AI and predictive monitoring. This keyword represents an area where the two topics are represented. Although machine learning is purely an algorithm being trained with abundant data, there is still a human component in the research surrounding this technology. If we ask where this data comes from, generally, the answer is through processes, experiences and performances, which plenty of them originated from human interaction.

SCOPUS' PERSPECTIVE ON THE LITERATURE DIFFERS FROM THE AUTHORS'

Keywords chosen and standardised by Scopus



On the contrary of the previous visualisation, where authors had the choice of keywords to describe their work, in this visualisation, the own search algorithms from the scientific database selected the more appropriate keywords to index papers. The way in which this is done is not publicly available, which also represents a problem with the transparency of AI.

This is an example of how, despite looking for the same, the internal search function indexes differently to the authors. But, why is it so different? If we compare, the most relevant keywords are humans, artificial intelligence and predictive monitoring, being the distance between them very significant. In general, this is a similar finding as the previous visualisation, but, this could change the narrative because, when we speak about humans, the related topics are complex, including social, political, economic and several other areas. Nevertheless, the complexity of interrelating different research areas is for granted; the question is, is a machine capable of sorting out and deciphering this complexity as a human would do? Would it be more or less optimal? The answer is that, as with every emerging technology, the growth of its applications tends to be unpredictable. Still, it is possible to investigate how to improve the foundation of those emerging technologies.

