

Report on the outcomes of a Virtual Mobility¹

Action number: CA19130

Grantee name: Piotr Wójcik

Virtual Mobility Details

Title: Exploring Research Visibility of the COST members: a Bibliometric Analysis of Topics

Start and end date: 09/10/2023 to 30/10/2023

Description of the work carried out during the VM

Description of the virtual collaboration and activities carried out during the VM, with focus on the work carried out by the grantee. Any deviations from the initial working plan shall also be described in this section.

(max. 500 words)

Within this virtual collaboration the team consisting of researchers representing five countries (Poland, Germany, Romania, Netherlands and Italy) measured the results and content of research published through the FinAI COST action 19130. In particular, we performed this analysis with respect to specific topics that relate to important deliverables of our action, i.e. “risk of using digital assets” (see deliverable 10), “stress tests for AI evaluation” (deliverable 11), and “finance failed trials” (deliverable 12).

We analysed all publications of the participants of the COST Action 19130 as for the end of March 2023. The publications were scrapped from the google scholar with the use of an R package “scholar”. Out of the total number of 216 Action participants for that moment, 51 did not have a google scholar profile. Therefore finally we took into account the publications of 165 unique researchers. We applied some additional filtering rules. As the Action focuses on the applications of artificial intelligence in finance including the risks related to digital assets, we limited our sample to publications not earlier than 2010, as 2009 is the year of the birth of bitcoin. In addition, as google scholar profiles might include not only research but also didactic materials (e.g. handouts for students) we limited our interest to the records that have non-empty journal name, which in case of google scholar profile also includes phrases like “working paper”, “proceedings”, etc.

To classify articles into topics we applied the state-of-art BERTopic algorithm (Grootendorst 2022), i.e. a transformer-based pre-trained language model based on the word embeddings. In a nutshell, embeddings are contextual representations of text. Embedded documents are represented in a vector space which enables comparing them semantically. Specifically, the most popular variants of the

¹ This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

embeddings are Generative Pre-trained Transformer (GPT) (Radford et al. 2018) and Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018). In accordance with its name, BERTopic algorithm uses a variant of the latter. Precisely, it applies Sentence Bidirectional Encoder Representations from Transformers (Sentence-BERT) (Reimers & Gurevych 2019). As the analyzed documents were written in different languages, we applied a multilingual variant of BERTopic. Parameters of the BERTopic algorithm were optimized using topic coherence measures, i.e., UCI (Newman et al. 2010), UMass (Mimno et al. 2012), UCI-NPMI (Aletras & Stevenson 2013). Cosine similarity which takes values between -1 and 1, was used for identification of topics' embeddings that were most similar to certain phrases: "risk of using digital assets", "stress tests for AI evaluation", "failed trials". In the first part of the analysis we focused on the titles of articles. In our database there were 8,066 unique records representing scientific publications from 2010 onwards with non-empty title. In addition, we extended the analysis on article descriptions (abstracts). In this part there were 5,485 unique records representing scientific publications with non-empty description (abstract) from the period 2010-2023.

Description of the VM main achievements and planned follow-up activities

Description and assessment of whether the VM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the VM. Agreed plans for future follow-up collaborations shall also be described in this section.

(max. 500 words)

The planned goals and expected outcomes of the VM were fully achieved. The results of our work were published on October 29th as an SSRN working paper and are available on the address:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4616662

The results of the analysis offer a comprehensive view on the portfolio of created articles and working papers and its topic classification, leading to important benefits for FinAI-related institutions across the European Union. The output of the VMG will be presented in one of the upcoming events of our COST Action. What is more, we are going to record a podcast with Vasile Alecsandru Strat, which will be published in open access. Therefore the Action participants will gain valuable insights into how established COST teams engaged in collaborative research can enhance their future publication output, how well the most important Action topics were already covered in their research and what are the other emerging topics that can lead to future joint projects proposals. Last, but not least, comprehensive analysis of the scientific output of the Action so far gives interesting insights for the last deliverable of the Action which is "an edited volume containing scientific achievements of the Action".