To the
Doctorate Commission
Al-Farabi Kazakh National University
Al-Farabi Avenue 71
Almaty 050040
Kazakistan

Reviewer's report for the Ph.D. Thesis titled "*The research and development of a module for an intelligent system for analyzing and evaluating the social mood of society in the media space of the Republic of Kazakhstan*" submitted by Karyukin Vladislav Igorevich for the Ph.D. degree in the specialty 6D070300 – Information systems

The thesis submitted by Vladislav Karyukin is devoted to the creation of automated systems for tracking user opinions, thoughts, and ideas about local and global events on Kazakh social media and news portals.

The main pain addressed by the thesis relates to the fact that the high volume of facts and opinions daily shared on social media makes such tracking impossible without automated methods. This advocates for effective and efficient analytics platforms to keep the pace with the explosion of published content. While a number of similar platforms already exist, many of them are not freely accessible, they primarily focus on resource-rich languages, and are optimized to operate on widely used social networks and news portals, leaving under-resourced languages and other social networks and portals poorly served. This is particularly evident for content published in the Kazakh language.

The goal of the thesis is therefore to develop a suitable technology for monitoring Kazakh news portals and social networks by extracting data from the internet, storing it in databases, performing quantitative and qualitative analyses of the content, and visualizing the results. To achieve this, the thesis introduces the Opinion Monitoring System (OMSystem), a new information system specifically designed for monitoring the social media space of Kazakhstan by analyzing content in both Kazakh and Russian. The OMSystem supports popular Kazakh news portals and social networks such as Facebook, VKontakte, Instagram, Twitter, and YouTube. At its core, it integrates various sentiment analysis methods that collect indicators about the level of activity in ongoing discussions in society, the level of interest in a particular topic, and the related social mood, in order to assess public sentiment and "social well-being." Additionally, it integrates marketing technologies to analyze the population's reaction to ongoing events and track the dynamics of user involvement in specific topics.

The thesis provides a complete overview of the work done towards the above objectives, by: i) discussing the initial study of strengths and weaknesses of current methods for monitoring and analyzing the social media space, ii) presenting the main modules and functionalities of the OMSystem architecture, iii) reporting on extensive experiments on sentiment analysis with different machine and deep learning algorithms including a BERT Transformer model for sentiment classification, iv) discussing the application of the developed technology to a specific use case (the monitoring of public opinion on the topic of vaccination against coronavirus infection), and v) introducing an application that, built on top of the OMSystem, analyzes its output to end up with visual representations of the social mood around political topics.

From a theoretical standpoint, the thesis is complete and solid. The applied methods and evaluation protocols are sound and one of the strengths of the work definitely lies in the thorough comparison of different machine learning techniques.

From a practical point of view, the actual development of a working platform, which has also been tested in a real-world scenario (and very topical, such as the COVID-19 emergency) is noteworthy: closing the loop from theory to practice is indeed quite rare for a PhD thesis. The caliber of the work presented makes the potential to "*allow timely political decisions to be made to ensure social balance and stability and promote political strategies in the internal and external environment*" (p. 8) a tangible possibility. Also commendable is the focus on under-resourced languages that, nowadays, represents an urgent need for users and a well-known challenge for the natural language processing community.

The good quality of the work is attested by the published scientific articles, which sum up to 12 peer-reviewed papers, including journal articles and one book chapter indexed in the Scopus database.

Summing up, the dissertation work of Vladislav Karyukin on the topic "The research and development of a module for an intelligent system for analyzing and evaluating the social mood of society in the media space of the Republic of Kazakhstan" meets the standard requirements imposed on a dissertation in this field, and should be accepted for the defense of Doctor of Philosophy (Ph.D.) in the specialty 6D070300 – Information systems.

Sincerely Yours,

Matteo Negri (Foreign Scientific Advisor)
Senior Researcher in the Machine Translation Unit
Fondazione Bruno Kessler (FBK)
Trento, Italy