2. Fake accounts detection approaches

2.2. "The Authenticity Matrix" tool

Nevertheless, "The Authenticity Matrix" has multiple limits. The main limitation of this tool lies in the fact that it allows research conducted on small volumes of data, as data collection and analysis must be performed manually. The second important limitation is that manual collection carries the risk of inaccuracy for the data collected due to human error. At the same time, data access was constrained by the use of personal Facebook accounts, and some shares were not visible due to platform restrictions. Additionally, the dynamic nature of social media, where posts, accounts, and interactions, can be deleted or blocked, poses challenges for research reliability and transparency. To minimize these limitations as much as possible, we implemented several measures, such as multiple cross-checks led by different team members on the same data set have been carried out to mitigate possible human errors.

We must also state that "The Authenticity Matrix" does not aim to classify the analyzed accounts into typologies such as bots, trolls, cyber, hybrid, etc. (see section 1.1), by establishing their automation or non-automation nature, but is limited only to classify the types of activity they exhibit in terms of authentic or inauthentic behavior.

5. Methodology

5.1. Dataset

To ensure the reliability and validity of the classification process, we employed a systematic approach to applying "The Authenticity Matrix" to our dataset. Four coders, all trained prior to the coding phase, were involved in the process. Their training included familiarization with the theoretical dimensions of the matrix and practical sessions focused on identifying and labeling account features in line with the coding framework. Two of the coders have previously worked with this tool on other corpus. We had a collaborative process that served as a calibration step, allowing the researchers to refine and align their interpretations of the matrix dimensions. Subsequently, the coders independently coded part of the corpus. This subset was used to assess inter-rater reliability. Krippendorff's alpha was calculated for the coded data, resulting in a value above 0.85. This level of agreement indicates a high degree of consistency between coders and supports the robustness of the classification procedure. The coding procedure specified criteria for each dimension of the matrix, each coding decision was guided by these criteria and any disagreements were resolved through discussion and consensus. This approach ensured both conceptual clarity and replicability, while minimizing subjective bias. Overall, the reliability

assessment and transparent documentation of the coding process reinforce the trustworthiness of the findings derived from the application of "The Authenticity Matrix".

We have expanded the methodological description to provide more detailed documentation of how coding decisions were made across the dimensions of "The Authenticity Matrix". Each dimension was operationalized through a set of predefined indicators and decision rules, which were discussed and agreed upon during the calibration phase. For instance:

- intended deception: was coded based on indicators such as the use of misleading accounts names or profile pictures,

- highly inauthentic behavior: referred to accounts that posted dozens of posts per day (187 posts/day, 184 posts/day, etc., see section 5.2.), or very frequent posts in short period of time (one post every 8 seconds, every 20 seconds, etc., see section 5.2.).