Enhancing Open-Source Intelligence: Introducing the Functional Intelligence Tool (FINT)

Laurens H.F. Müter*
Netherlands Police
Utrecht University, Nationaal Politielab AI
Utrecht, the Netherlands
l.h.f.muter@uu.nl
*Corresponding author

Jacco van der Plaat
Utrecht University
Utrecht, the Netherlands

Anna Zoet
Netherlands Police
Utrecht, the Netherlands

Christof van Nimwegen
Utrecht University, Nationaal Politielab AI
Utrecht, the Netherlands

Remco C. Veltkamp
Utrecht University, Nationaal Politielab AI
Utrecht, the Netherlands

2024 9th International Conference on Big Data Analytics (ICBDA) | 979-8-3503-5296-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICBDA61153.2024.10607337

***Abstract*— With an increasing number of protests involving a more diverse group of people, the Dutch Police have to analyse a large number of messages posted on social media platforms, which becomes increasingly difficult to scrutinise for the Open Source Intelligence (OSINT) analysts of the Dutch Police. This is not only because of the increased volumes and velocity at which messages are sent, but also due to the diversity of language usage and the amount of bias that hides the crucial information. To shed light on visualising social media posts, this study combines user-centric design with ecological interface design principles to create the Functional Intelligence Tool (FINT). This tool is tested on twenty participants, of whom four OSINT analysts of the Dutch Police. The tool is reported to be consistent, well-integrated and easy to learn, even for participants without a background in OSINT analysis.**

Keywords—protests, data visualisations, big data, social media

I. Introduction

Protests are a growing phenomenon worldwide. Ortiz et al. [1] show that the number of protests has steadily increased since 2006. Furthermore, protests have become more politically oriented due to disappointment with malfunctioning democracies, politicians, and an overall lack of trust in governments. This increasing number of protests has reached the Netherlands, where politicians struggle with large-scale protests and ever-increasing occurrences of rioting behaviour [2]. The Dutch Police reports that violence against the police remains high, with verbal threatening behaviour towards police officers escalating towards open violence and even murder attempts [3], which is referred to as public disorder in this paper. To prevent excesses of public disorder, the police need insights into social unrest such as rioting behaviour. Therefore, the Dutch Police use open-source intelligence (OSINT) to search open sources for indicators of violence in upcoming protests. Lande, Subach and Puchkov [4] consider OSINT as a key tool in modern cybersecurity. It comprises the search, selection, collection and analysis of intelligence information from sources that are open to everyone. One source of intelligence information is social media. Notably, OSINT is used for social media monitoring [4] as a form of decision support [5].

Although OSINT analysts receive specialized training, the large volumes of data are an enduring problem. Moreover, the increasing social media usage will lead to even more data in the future [6]. However, this increase in data also leads to more personal and useful information, as social media users share things voluntarily, even their dissatisfactions [7]. With the rise of violent and rioting behaviour during protests, the need for tools to analyse large quantities of social media messages increases correspondingly.

Although it is tempting to monitor many social media accounts, there are strict regulations in place that prohibit the use of large-scale social media analysis without proper authorization and adherence to privacy laws. For example, the General Data Protection Regulation (GDPR) states that the Posts should not lead to the identification of individual users [7]. To ensure Privacy by Design (PbD), systems can monitor and analyse Posts automatically with the necessary privacy protection, thus releasing pressure on OSINT analysts in their line of duty. To contribute to this idea, this study aims to provide an open-source intelligence tool to aid the Dutch Police in their task of scrutinizing social media to prevent violent behaviour at large social gatherings, while respecting privacy requirements stated by the GDPR. To concretize the aim of this study, the Functional Intelligence Tool (FINT) is developed using Artificial Intelligence (AI) to monitor a stream of Posts related to protests. When indications of upcoming violence are detected, FINT has to provide further understandable details. This leads to the main research question of this study:

How to visualise relevant Post content in the context of predicting violence for OSINT?

This question is divided into four subquestions. Because OSINT analysts with different backgrounds, experiences and ages will work with FINT, the tool should be useful for a wide range of people. This leads to the first subquestion *S1: How do the more experienced OSINT analysts compare to the less* *experienced analysts when using FINT?* Next, FINT is compared with tools that are currently used by OSINT to find out if there are improvements in quality, through *S2: Does FINT demonstrate a more effective presentation of social media posts than current tools?* Furthermore, the usability of the current tools and FINT are compared, which led to *S3: Is FINT more user-friendly than existing tooling?* For FINT to be introduced swiftly within the team, a key concept in its usefulness is the integration of existing tools and processes. This led to the fourth hypothesis *S4: Can FINT be seamlessly integrated as an extension of current tools?*

During this research, a bare-bones implementation of FINT is created to gain insights into the research question. The barebones FINT is then tested in a working environment with a group of students and four OSINT analysts of the Dutch Police, using a set of synthesized Posts. With this setup, a first impression of the performance of FINT is captured, which, in turn, will serve as a baseline for further development. Another advantage of this setup is that OSINT will obtain valuable insights into the possibilities of using AI in their workflow with minimal effort.

With this research, a practical contribution is made to the OSINT work field of the Dutch Police. In addition to the practical use of the FINT dashboard, a scientific contribution is made to the application of existing machine learning models in real-world scenarios using an approach where ecological interface design integrates with user-centric design. The resulting dashboard is easy to use and could serve as a foundation for future research. As supplemental datasets are made publicly available, future research will benefit from this study as a baseline. Finally, different methods of visualising Posts are discussed and tested within a practical context, providing new insights into the usage of these visualisations.

II. Related Work

There are various tools available to monitor and visualise sentiment on platform X (formally Twitter). These tools can help to gain a quick overview of how people react to certain large-scale social events and how these reactions evolve. Sharifi, Hutton, and Kalita [8] explored a text-based summarization approach in their research. In a similar vein, Marcus et al. [9] introduced TwitInfo, a tool equipped with visualisations designed to unveil patterns and trends in sentiments. Notably, many tools tend to categorize emotions into binary positive or negative expressions [9]–[11]. However, some researchers have delved into a broader spectrum of emotions. For instance, Kempter et al. [12] developed EmotionWatch, a tool capable of expressing twenty emotions, including but not limited to anger, worry, guilt, and sadness. Beyond sentiment analysis, researchers have also directed their efforts toward applications in business intelligence. Sijtsma, Qvarfordt, and Chen [13] introduced Tweetviz, a visualisation tool tailored for business intelligence purposes.

A. Commercial Social Media Monitoring Tools

There are many commercially available tools to monitor social media, each serving different purposes, a top ten related tools are provided:

- **Buffer**[1] is a social media scheduling tool that offers basic analytics and reporting features for monitoring social media activity.
- **Brand24**[2] is a social media monitoring and analytics tool that helps businesses and individuals track their brand mentions and monitor online conversations relevant to their brand or keywords.
- **Brandwatch**[3] is a social media listening and analytics tool for businesses to gain consumer insights and track brand mentions across various social media platforms.
- Hootsuite[4] allows users to manage multiple social media accounts, schedule Posts, and track social media performance. It also provides basic monitoring features.
- **Hubspot**[5] contains a customer platform made up of six core products for marketing, sales, customer service, content management, operations, and commerce. The marketing service includes social media analytics.
- Keyhole[6] is a hashtag tracking and social media analytics tool that provides insights into brand mentions, trending topics, and campaign performance.
- **Meltwater**[7] is a media intelligence and social media monitoring platform that helps businesses monitor, analyse, and engage with online content. It provides a range of features for social media analysis, reputation management, and media monitoring.
- **SocialPilot**[8] is a social media management tool that is designed to help individuals and businesses manage their social media accounts more efficiently. It provides features for scheduling Posts, analysing performance, and collaborating with team members.
- **Sprout Social**[9] is a social media management and analytics tool that provides monitoring, engagement, and reporting features for multiple social networks.
- **Talkwalker**[10] is a social media analytics and monitoring tool that offers real-time brand mentions tracking, sentiment analysis, and competitive intelligence.

The practical context of OSINT requires a tool to have different features which might help to make estimates of violent behaviour and gain insights into popular topics related to upcoming protests. Based on the available textual information from Posts, the most commonly used features are summed in table I.

[1] https://buffer.com
[2] https://brand24.com
[3] https://www.brandwatch.com
[4] https://www.hootsuite.com
[5] https://www.hubspot.com
[6] https://keyhole.co
[7] https://www.meltwater.com
[8] https://www.socialpilot.co
[9] https://sproutsocial.com
[10] https://www.talkwalker.com

The most common feature requirements for OSINT include:

- **Topic** clustering (TC) show relations between topics
- **Hashtags** (HA) show most popular phrases
- **Emojis** (EM) show tone of the messages through emojis
- **Sentiment** (SE) estimate the probability of disturbances during upcoming protests
- **Filtering** (FI) filter out specific terms
- **Raw Posts** (RA) gain context extended to other features
- **Geo-locations** (GE) estimate the location of possible riots

These features are rated for every tool listed on a five-point scale; - - (not included), - (included, not useful for OSINT), 0 (included, might be useful for OSINT), + (included, useful for OSINT), ++ (included, very useful for OSINT).

Given that the primary focus of these tools revolves around assisting businesses in evaluating the effectiveness of their campaigns, particularly for endeavours like introducing new products, their features are predominantly tailored to the purpose of monitoring how a group of individuals responds to particular topics. Consequently, these tools may be less practical for monitoring broader subjects such as social unrest or group behaviour in the context of an upcoming protest. Furthermore, it's noteworthy that certain social media platforms, including X, are reluctant to have governments utilize their services to extract information regarding public sentiment.

TABLE I. COMPARISONS BETWEEN SOCIAL MEDIA ANALYSIS TOOLS ON FEATURES RELEVANT FOR OSINT

Tool Name	TC	HA	EM	SE	FI	RA	GE
Buffer	- -	-	- -	- -	- -	- -	- -
Brand24	0	+	0	0	+	+	0
Brandwatch	+	0	0	0	+	++	+
Hootsuite	- -	+	- -	- -	+	+	+
HubSpot	- -	- -	- -	- -	- -	- -	- -
Keyhole	0	+	0	0	+	0	+
Meltwater	+	+	+	0	++	++	+
SocialPilot	- -	- -	- -	- -	+	- -	- -
Sprout Social	0	++	- -	0	+	+	+
Talkwalker	+	++	+	0	++	+	+

B. *Current OSINT Workflow*

How OSINT analysts use their tooling is described by Muter et al. [14] who interviewed two members of the OSINT team from the Dutch Police. The OSINT team mainly focuses on obtaining a global view of large-scale social events and expressions of discomfort toward political spokesmen to safeguard public speakers and prevent violent behaviour.

Currently, OSINT uses commercially available tools that use temporal changes in the number of Posts and RePosts to identify popular topics. Contemporary OSINT tools feature visual representations such as word clouds, hashtag clouds, emoji clouds, and Posts identified as highly influential based on metrics such as likes, followers, and RePosts [14]. Samples of text in Posts enriched with simple sentiment scores are included. However, most insights are gathered from scrutinizing the tone of language from messages related to specific topics and the intuition of the analyst.

Insights resulting from the information-gathering process are often presented on a single page without technical terms to be understandable for decision-makers, as they often must act fast with very limited margins of error [14].

C. *Machine Learning Models*

Multiple models are constructed and tested within the context of OSINT by Muter et al. [14], which serves as a starting point for the development of the FINT dashboard. The most useful models include:

- **Networks of hashtags** are constructed by linking hashtags that coexist in the same Post. The linked hashtags can be enriched with Latent Dirichlet Allocation (LDA) to gain insight into the most frequently used hashtag.
- **Sentiment** can be beneficial when the model performance is good. However, when the model does not perform well, sentiment scores can easily become misleading as it often states something about a bulk of messages.
- **Emojis** can visualise the most common feelings among groups. For example, farmers use tractors and the Dutch flag in their messages.
- **Geo-locations** which include the geo-tags when available from Posts. Although only a small portion of Posts contains geo-tags, they can provide information on where most activity takes place.

Since the hashtags network appears to be useful, this network is extended by including all descriptive words from Posts with LDA. However, since a Post often contains more descriptive words than hashtags, an additional clustering is made on the topics to get a more thematic view. This way the analysts can determine an overlapping theme by looking at the descriptive words within a cluster. Another combination of models is constructed using the sentiment scores of the Posts with the hashtags of the Posts, these hashtags are then sorted on the most negative and most positive sentiment.

D. *Design Science Methodology*

Design Science comprises the study of iterating design processes to improve them [15]. Design Science Research (DSR) seeks to generate knowledge of scientific validity and offers practical utility. The field of Design Science is defined in multiple ways [16]–[18]. Artefacts produced by Design Science, 'Design Theories', often refer to models, frameworks or bodies of knowledge. These artefacts provide information on the architecture of Information Systems and can be seen as design principles. However, there has been little research as to how such Design Theories could be communicated, justified or developed [18]. How to apply DSR correctly remains cumbersome, even though there are different approaches.

1) First Approach: User-Centric Design: User-centric design (UCD) is often used within Design Science to develop new tools. As the name implies, when using this approach, the user is centred during the design, resulting in an artefact based on the user's needs. Saffer suggests four approaches [19]; Activity-Centred Design (1) which focuses on tasks and activities that need to be fulfilled, System Design which focuses on system components, Genius Design (2) which includes the ability of the designer used to make products, and User-Centred Design (3) which focuses on the needs and user goals. Six principles of UCD are described by Chammas et al. [20], and can be summed as:

- **Explicit Understanding of Users:** Base the project on a clear understanding of users, their tasks, and their environments. Consider the context of use, as it is crucial for establishing system requirements.
- **User Involvement Throughout Development:** Engage users throughout the development process to gain valuable insights into the context of use. The level of engagement should be tailored to the specific project.
- **Progressive Assessment:** Conduct and refine the project through assessments focusing on the user to minimize the risk of the system not meeting user needs and desires. Progressive assessment identifies critical points for effective product acceptance.
- **Iterative Design Process:** Iteration Design involves reviewing design specifications based on new information acquisition, aiming to develop a system that will meet requirements and user expectations.
- **Holistic User Experience Design:** Address the entire user experience, considering the presentation, functionality, system performance, interaction behaviour and assistive capabilities of the interactive system. Take into account user attitudes, skills, habits, and personality, as well as their previous experiences, capabilities, limitations, preferences and expectations.
- **Multidisciplinary Project Team:** Assemble a project team with multidisciplinary skills and perspectives. Include members from diverse areas to share skills, experiences, and views, benefiting the project with a range of perspectives and expertise.

2) Second Approach: Ecological Interface Design: Ecological Interface Design (EID) offers an alternative to UCD, incorporating the user's surroundings as a contextual element in dashboard development. EID was introduced by Rasmussen and Vicente in 1989 [21] for process control, to enhance the interactions between humans and automated systems by creating interfaces that mirror the deep structure of the work domain, supporting human problem-solving activities based on skills, rules, and knowledge (SRK) [21].

In contrast to technology which emphasizes either the human or technological aspects, EID initiates the design process by concentrating on the work domain, referred to as the "ecology" [22]. This approach facilitates and defines purposeful patterns of activity, irrespective of whether they are performed by human or automated agents [22]. This makes EID suitable for applications in various domains, including process control [23], health care [24], command and control [25], and aviation [26], [27].

Specific components of EID include the abstraction hierarchy for work domain analysis and the SRK model of human cognition, as described by Rasmussen et al. [21]:

- **The Abstraction Hierarchy (AH)** offers valuable insights into functional interactions, enabling the effective partitioning of functional complexity in work domains. This facilitates the construction of representations that capture the 'deep structure' or semantics of the work domain, providing a basis for organizing information in both internal and external representations.
- **Cognitive Work Analysis (CWA) framework** provides a systematic approach to understanding the cognitive work from a given domain. By analysing the work domain, control tasks (ConTa), strategies, social aspects and personal competence, designers can develop interfaces that align with the deep structure of the work domain.
- **Control Task Analysis (ConTA)** is part of the CWA framework and focuses on the examination and understanding of the control tasks performed within a given work domain. The primary objectives of ConTA include; Identifying Control Tasks (1), Decisions and Actions (2), Temporal Aspects (3), and Feedback Mechanisms (4).
- **Worker Competencies Analysis (WCA)** uses the SRK model which sheds light on the coupling between cognitive agents and control problems. The skill-based level acknowledges the efficacy of direct manipulation and perception for control processes with small time constants. As effective time constants increase, the rule-based level recognizes the need for more pre-packaged responses to triggers. Given the complexity of the domain, operators often employ formal (knowledge-based) reasoning to diagnose problems and discover unforeseen opportunities not considered in the original system design.

3) Third Approach: Combining UCD and EID: A human factors method that combines EID and UCD is constructed by Revell et al. [28]. This method is named UCEID and is useful for complex sociotechnical systems. Since UCEID ensures user needs are appropriately represented in the constraints-based models provided by the CWA framework. Although this work looks promising, there is no study available yet that tested UCEID within a practical setting. Moreover, many projects focus on either EID or UCD and often there seems no need to combine the two.

III. Dashboard Design

Multiple steps are taken to set up a tool to support the OSINT team in gathering information from online social media platforms. First, an initial interview was conducted to gather insights into OSINT-specific problems when using current tools. The findings from the initial interview were used for a user-centred design approach, extended with EID components, to fit within the ecology of the OSINT team.

A. Requirements and Constraints

Although OSINT was very open about its processes, there were still questions that could not be answered for safety reasons. In particular, the OSINT team could not disclose any information on current tools, since this would make them vulnerable to hackers. The OSINT team also was unable to share anything about ongoing investigations or internal information requests, for any information leak could benefit criminals. For example, if criminals become aware of being traced, they might destroy evidence to complicate the process of tracking and prosecuting them. For the same reason, OSINT analysts are not allowed to discuss any completed information requests. Moreover, any access from archived files is blocked, which means even analysts who conducted the research are denied access to previous projects. Although this strict process could disturb many possibilities to learn from previous mistakes, the stakes are too high to ignore any safety precautions.

OSINT analysts did share information about their processes and the kinds of information they are looking for when preparing for large-scale events in general, even though there is little room for content-related information. Combined with a quick search on currently available tooling concerning social media, their answers gave a good impression of what an information request might look like.

In addition to their workflow interview, OSINT analysts provided valuable insights into their approach to analysing a non-existing event based on synthesized messages, as detailed by Muter et al. [14]. Furthermore, OSINT analysts shared information about their technical skills and backgrounds, enabling the incorporation of external factors into the design process, as discussed in the context of UCEID.

Permitted information sources include topics, emojis, hashtags and location data. Other sources, such as usernames, mentions, and any identification of individuals are prohibited by the GDPR [29]. Analysts are also restricted from tracking users, entering non-public groups without permission, linking data from different profiles, acquiring more information than necessary for solving the research question and storing raw information. Additionally, the use of information from previous events, such as filtering individuals or groups that caused problems in prior occurrences, is not allowed, even though the information was originally public. Moreover, the usage of any data that reveals personal details is prohibited [29]. By default, all information is blocked once an event has ended, although the experiences of the analysts are retained. These aspects are considered during the design process of FINT.

B. Design Approach

A simplified version of the UCEID from Revell et al. [28] is used to develop FINT. Figure 1 shows the study stages, as based on the design cycle of Wieringa. This cycle is used for design processes when doing Design Science Research as described in Section II-D and comprises of the stages: problem investigation, treatment design and treatment validation. Note that 'treatment' refers to a solution of the defined problem. It is a subcycle of the overarching engineering cycle, which includes treatment implementation and implementation evaluation [30].

- **Problem Investigation** contains three steps. Based on previous studies, a problem definition (1) is constructed with a clear context. Additional context on existing processes is gathered within an initial interview (2) with the OSINT team. Finally, existing tools are scrutinized against the gathered requirements (3).
- **Treatment Design** contains four steps which are represented within a scenario (1) of synthetic Posts about four unrelated events (about politics, climate, and leisure). Data, extracted from the user interviews, is used for thematic analysis (2) to consolidate and triangulate themes. An abstraction hierarchy (AH) (3) is then presented as a diagram with five levels of abstraction (purpose, value, purpose-related functions, functional purpose, physical object), a control task analysis (ConTA) is performed (4), and a worker competencies analysis (WCA) is used to identify the cognitive skills required to perform OSINT-related tasks (5).
- **Treatment Validation and Implementation** includes two steps: validation of artefacts (1) to decide on which requirements and artefacts should be included in FINT and to implement FINT (2).
- **Treatment Evaluation** takes three steps: FINT is tested within a controlled environment of the OSINT team, using the scenario from the first step of the treatment design (1). Next, an evaluation interview is conducted with the OSINT analysts to obtain their experiences with FINT (2). Finally, FINT is also tested by people who do not have any experience with OSINT (3).

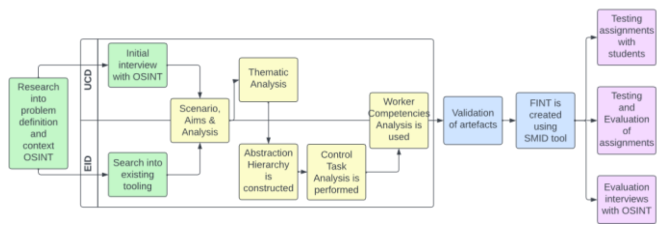

Fig. 1. Different stages of the study following the design cycle of Wieringa; problem investigation (green), treatment design (yellow), treatment validation and implementation (blue), and treatment evaluation (purple).

C. Design System

Little literature on design systems is available, several attempts have been made to define a Design System (DS) [31], [32]. A DS is a collection of reusable components to enhance the assembly of an interactive website or application. The most common DS's are Bootstrap, Foundation, and Materialize, which differ on customizability, speed of development, and dependencies [33]. Bootstrap or Materialize are most common, although Shenoy & Prabhu [34] argue that smaller design systems such as Skeleton, Milligram, and UIkit might be preferable due to performance and ease of use. After consideration, a final decision was made to utilize Materialize, since it is widely available and already used by the Dutch Police during other projects.

FINT is built upon the existing tool 'SMID', which was previously constructed to facilitate the emergency control centre of the Dutch Police [35]. This tool included different visual elements and required access to a real-time stream of Posts. However, for this experiment, a predetermined set of Posts was used to ensure reproducible results. These Posts are accessible through a REST server, linked to a VUE front-end.

D. Resulting Dashboard (FINT)

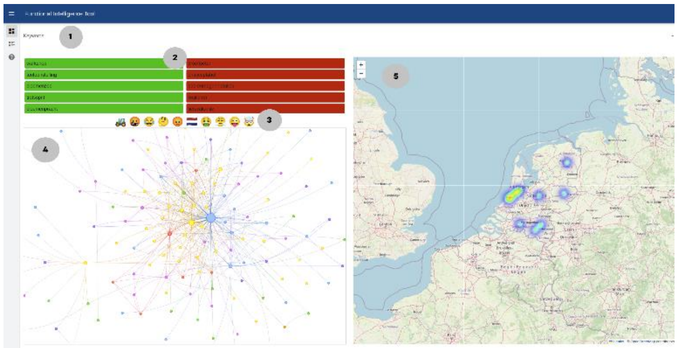

Fig. 2. FINT main dashboard with dummy data. Multiple visualisations are shown (1) searchbar, (2) bookmarks with sentiment, (3) emojis, (4) topics network, and (5) geo references shown on a map.

The resulting dashboard contains six elements that are all based on different views of the data. The **searchbar** (1) allows multiple terms to be inserted and utilizes an AND filter to find messages. This search is then updated automatically in all five other views. The **sentiment scored bookmarks** (2) represent hashtags from the messages which are coloured based on the sentiment score of the surrounding message. The **emojisbar** (3) represents a tf-idf representation of emojis available in the filtered messages, displaying the most frequently used emojis on the left. The **topicsnetwork** (4) contains topics extracted from the messages using Python's Spacy library. These topics are also presented as tf-idf with dots ranging in size, most frequent terms are enlarged. The terms are connected when they occur in the same message. The network of terms and connections is clustered using the KMeans clustering method of the SKlearn Python library. The resulting network clusters are presented in eight distinct colours. The **map** (5) contains geo-tags from locations that are being discussed within the messages. Finally, a list of **raw messages** (6) is presented in another tab so the user can read some of the messages for additional context.

Visualising these models resulted in the main dashboard as shown in Figure 2 which contains the search bar, sentiment bookmarks, emojis, topic network, and the map with geotags. The list of raw messages is contained within a dedicated tab (accessible from the left-hand side menu), shown in Figure 3. This page also contains a help section, as depicted in Figure 4, to serve as a quick guide for users to learn what different visual elements indicate.

In future iterations, we plan to explore other information streams, such as those from other social media platforms or news media, which can also be integrated through REST APIs.

IV. Method

The main research question - *How to visualise relevant Post content in the context of predicting violence for OSINT* - is divided into four sub-questions which form the basis of four distinct hypotheses. **S1**, on how more experienced analysts compare to less experienced analysts leads to hypothesis ***H1****: OSINT experts yield superior results compared to junior analysts and students*. The second sub-question **S2** on how FINT compares to tools currently used by OSINT, is transformed into the second hypothesis: ***H2*** *FINT demonstrates a more effective presentation of social media posts than current tools*. The third sub-question **S3** considers how user-friendly FINT is, corresponding to the hypothesis ***H3****: FINT is more user-friendly than existing tooling*. Finally, sub-question four **S4** on the integration of FINT in the current work processes are transformed into ***H4*** *FINT can be seamlessly integrated as an extension of current tools*.

TABLE II. Group Characteristics of Participants

Age Range	Num. Participants
< 20	6
20 - 30	2
30 - 40	6
40 - 50	0
> 50	2

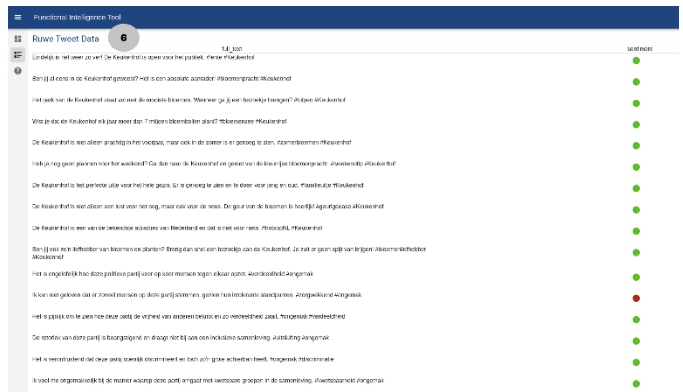

Fig. 3. The second tab of FINT contains the raw Posts. Filterings applied in the main dashboard are adapted within this view.

Fig. 4. The third tab shows a help page with all visual elements from the main dashboard with a short description and an explanation on how the visual element can be used.

A. Participants

This study involved twenty participants, four of whom are working as OSINT analysts for the Dutch Police. Two analysts have between fifteen and twenty years of work experience, whereas the other two have less than five years of experience.

The less experienced OSINT analysts range in age from twenty to forty years, while the experienced analysts range between forty and fifty. Three of the OSINT analysts specialize in operational roles, one serves as a generalist intelligence analyst. The education level of the OSINT analysts is comparable to an applied science level of education (HBO in Dutch), although OSINT is not a formal applied science degree.

The age range among non-OSINT participants varies, with the youngest individuals below twenty years and the oldest participants exceeding fifty, as detailed in Table II. The non-analyst participants encompass a range of qualifications. One participant holds a senior secondary vocational education degree (MBO in Dutch), three participants have an applied sciences level (HBO in Dutch), and twelve participants are studying or have achieved a degree at the academic university level (WO in Dutch), see Figure 5. Three participants reported having a pre-university secondary education degree (VWO in Dutch) as their educational background. Five participants are currently pursuing a bachelor's degree in biomedical science. One participant has a background in mathematics and physics, another in cultural anthropology, one in psychology, one in human resource management, one in cultural sociology, one in nursing and social work, one in technology and mechatronics, and one participant specialises in professional organizing.

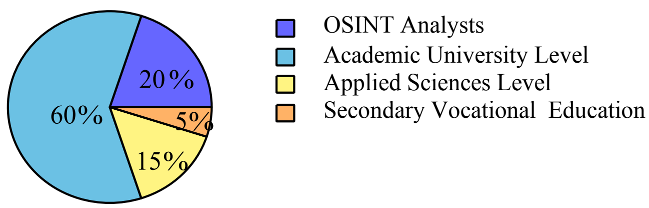

Fig. 5. Educational Background of all attending participants. The educational background of the OSINT analysts is presented separately since they undergo education within the police academy, which does not formally align with the standard educational systems in the Netherlands.

B. *Experimental Setup*

The questionnaires are created using Qualtrics [11]. All questions and most of the synthesized Posts are in Dutch. The first questions relate to the participant and include age, years of experience, background, and educational level. After a short introduction to FINT, the participant receives a task that should be fulfilled using FINT. Finally, a Dutch System Usability Scale (SUS) questionnaire [36] is presented where the user can fill in his experiences with FINT.

1) Experience with Existing Tooling: To establish a baseline for the evaluation of current social media monitoring tools, a questionnaire is administered. Using the Dutch SUS questionnaire, the usability is measured, providing a foundational benchmark.

2) Introduction to FINT: Following the initial SUS questionnaire, participants receive a brief introduction to FINT, elucidating the new tooling features present in the dashboard. This session allows participants to pose questions and receive a written manual. This introduction aims to familiarize participants with FINT for effective use in subsequent steps.

3) FINT Assignment: After becoming familiar with the functionalities of FINT, participants start with a task to extract information from a collection of synthesized Posts of a fictional scenario. The assignment includes four questions, each carrying a maximum of 25 points, allowing the participant to earn a total score of 100 points. These scores serve as a basis for evaluating the performance of OSINT experts in comparison to junior analysts and students. The four questions pertain to the content of the synthesized Posts, encompassing: (1) the identification of themes and topics within the Posts, (2) the assessment of which topics and themes have the potential to incite public disorder, (3) an estimation of the likelihood of individuals engaging in protests, and (4) an identification of the places and locations where protests are likely to occur.

4) Usability of FINT: After the assignments, the participants filled in a Dutch SUS questionnaire to evaluate their experience and FINT's effectiveness compared to other tools. The SUS questionnaire is presented using a Dutch Likert scale. Additional questions are included to focus on specific features: (1) the helpfulness of the search functionality, (2) the utility of sentiment bookmarks, (3) the usefulness of emojis in the interface, (4) the effectiveness of the topics network feature, (5) the helpfulness of geolocation tools and (6) the usefulness of raw Posts in the FINT interface. These responses aim to gather insights into the usability, providing valuable information on FINT's usability and user satisfaction.

5) Exist interview: Finally, the OSINT analysts are interviewed to explain which features they liked, how FINT might be useful practice, and what suggestions they have.

C. *Measurements*

The following information is recorded during the experiment to make comparisons between OSINT experts and students concerning FINT versus existing tooling:

- **Usability questionnaire** with additional questions on the elements from FINT.
- **Number of correct answers** the score is calculated by taking the sum of scored answers.
- **Time** of interaction with the tool, starting when the first question is presented, with a maximum of 20 minutes.
- **Exit interview** additional questions on how certain the participant is regarding their work.

V. Results

The results of the measurements are reported per hypothesis.

H1 is not supported by the results. This shows that there was no significant difference between experienced OSINT and less experienced analysts. Moreover, the more experienced OSINT analysts scored lower on the assignment, with an average score of 60.50% in comparison to 77.50% for the less experienced OSINT analysts. However, experienced analysts have used combinations of multiple visualisations more often, which made it easier for them to answer the final question correctly. For example, one analyst reported using the locations indicated on the map to acquire search queries to find Posts with a location indication.

There was also no significant difference found between the OSINT analysts and the non-OSINT analysts participants,

[11] https://survey.uu.nl/survey-builder/

although the average scores of the OSINT analysts were higher than the latter, with an average score of 67.25% (SD = 12.79) versus 61.44% (SD = 9.85). In particular, OSINT analysts scored better on the question about assessing themes that indicate social unrest or public disorder, even though no significant difference was found for individual questions. In general, the non-OSINT participants scored best on the first and final questions (identify themes and identify location respectively), whereas OSINT analysts scored best on the second question.

The time it took for the OSINT analysts to complete the tasks was slightly longer than for the non-OSINT participants. However, this difference was insignificant. The overall time participants spent on the first question was significantly longer than the other questions ($t(62) = 6.23$ with a $p < 0.0001$).

The second hypothesis, ***H2***, was also not supported by the results. This means that FINT does not demonstrate a more effective presentation of social media posts than the current tool, since no significant differences were found between the two. However, the statement that FINT contains features that are not present in current tooling scored marginally significant with a p-value of 0.09 on a one-tailed t-test. Some of the features were reported to be useful in solving the questions of the assignment, such as the search bar with a significant result on a one-tailed t-test $t(6) = 3.351$ with a p-value of 0.0077, and the topic network $t(6) = 2.34$ with a p-value of 0.02 also on a one-tailed t-test. The sentiment representations in FINT were reported to be more useful than sentiment scores in the current tool, but this was not significant.

H3 did not reveal significant differences, so FINT is not more user-friendly than existing tooling. Nevertheless, certain distinctions favoured FINT, such as a higher score (0.75) in ease of use compared to the current tooling (0.5). Additionally, FINT received a better score (0.75) in terms of user learnability compared to the existing tools (0.5). Analysts reported minimal learning requirements for FINT (average score of 1.0), whereas the current tooling scored 0.0. This difference was marginally significant ($t(6) = 1.51$, p-value = 0.09) in a one-tailed paired t-test. Furthermore, analysts found FINT to be less complex than their current tool (average score of 1.0 versus 0.75). However, confidence levels in using FINT were lower among analysts (0.5) compared to their current tool (0.75).

No notable disparity in overall SUS scores was observed between OSINT analysts and other participants. FINT did receive a significantly positive SUS score from both OSINT analysts ($t(6) = 1.95$, p-value = 0.0497) and non-OSINT participants ($t(30) = 3.87$, p-value = 0.00027). Especially items such as FINT not being complicated, well-integrated, consistent, easy to learn, and not difficult to use.

However, non-OSINT participants expressed lower confidence in using FINT compared to OSINT analysts ($t(18) = 1.99$, $p = 0.31$), even though non-OSINT participants considered more visualisations to solve the assignment. Sentiment Bookmarks were considered to be significantly more helpful for the non-OSINT participants than for the OSINT analysts ($t(18) = 1.75$, p-value = 0.048). The map feature was also reported to be more helpful by non-OSINT participants in contrast to the OSINT analysts ($t(18) = 2.245$, p-value = 0.0188), as were raw Posts ($t(18) = 1.85$, p-value = 0.04).

Both OSINT analysts and non-OSINT participants concurred that emojis did not contribute significantly to solving the assignment. Non-OSINT participants rated the usefulness of emojis significantly lower than the map ($t(30) = 2.22$, $p = 0.017$) and raw Posts ($t(30) = 2.29$, $p = 0.015$).

H4 showed no significant results, although the integration of FINT as an extension of current tools seemed to be received positively. During the exit interview, analysts stated that FINT felt like an oversimplification, but it was hard to pinpoint what was missing exactly. Intuitively, OSINT analysts would seek value in statistics regarding the number of Posts and how topics would change over time, but they found it difficult to say how these features would improve the visualisations of FINT. The main features of FINT and their usefulness for the OSINT workflow were discussed during the exit interview:

- **Search Bar** was reported to be useful for navigating through the messages. The filtering on multiple terms was considered limited, whereas the OSINT analysts staged that including OR filters might increase its usability.
- **Sentiment Bookmark** Although sentiment bookmarks were not often used to solve the assignments, it was reported to be useful to have information about the sentiment of a message, and for example filter on sentiment values. The sentiment scores added value within a practical setting.
- **Emojis** are already present in current tooling and can yield additional information on crowd sentiment. However, emojis do not provide much insight in general.
- **Topic Network** was considered as a useful feature. By connecting multiple topics, analysts could get a feeling of what was happening in the dataset. However, they reported room for improvement, such as simplifying the representation by only showing the clusters and connections between them and allowing users to 'zoom in' on a cluster to reveal underlying topics. This way, the user is less distracted by the large number of dots on the start screen and can focus more on the overall picture.
- **Geo Map** is already present in current tooling. In contrast, the FINT representation was considered less useful than the current tooling, which displays the number of messages per region on the map.
- **Raw Posts** showed a neutral response on usefulness. However, one of the analysts used the browser's Ctrl-F function on raw Posts to specifically search for 'protest'. The number of matches then served as support for the assumption that a protest would occur.

VI. Concluding Remarks

There was no significant difference in scores between the OSINT analysts and the non-OSINT participants (**S1**), reinforcing the assertion that FINT is user-friendly and easily accessible even for users without a background in OSINT. Although OSINT analysts stated that FINT contains features that are not present in tools currently used, such as topic networks and sentiment bookmarks, the results did not yield a significant difference from current tooling (**S2**). In comparison to existing tooling, FINT performs well on ease of use and learnability, and the tool is reported to be less complex and more approachable than existing tooling. Further support for FINT's user-friendliness emerged from the SUS questionnaire, where both OSINT analysts and non-OSINT participants reported positive experiences with the tool. Although the SUS questionnaire was positive for FINT, it did yield a significant improvement over the current tooling (**S3**). FINT can be seamlessly integrated as an extension of current tooling based on a marginally significant result from the questionnaire and the exit interview with OSINT analysts. The search bar of the existing tooling is more extensive and might be useful to filter the data before it is led into FINT. More advanced features, such as sentiment bookmarks and a topic network, might add value to the extension of current tooling (**S4**).

An explanation for the lack of significant differences in assignment scores could relate to the questions being too straightforward and not fully capturing the complexity of typical OSINT tasks. Given the experimental nature of this setting, replicating a real-time scenario involving the analysis of hundreds of thousands of Posts is inherently challenging. Nonetheless, the OSINT analysts, as revealed in exit interviews, considered the simplified scenario presented in this experiment to be realistic. Future research might reveal how FINT performs on real-world data within a practical setting.

Another limitation is the use of synthetic Posts in this study, which might yield different results compared to real-world Posts. To address this, the main author amassed sets of real-world Posts through previous research projects and supervised teams in labelling these messages. This experience is used to ensure alignment with authentic data experiences. Additionally, leading up to this study, the main author engaged in bi-weekly discussions with OSINT experts (not involved in the study), where real-world scenarios and Posts were thoroughly examined.

On average, the less experienced OSINT analysts obtained a higher score on the assignment than the experienced OSINT analysts. More specifically, the answers of the experienced OSINT analysts contain details and assumptions that were not present in the dataset while the less experienced analysts did describe the present data in more detail. For example, analysts reported elections as a theme while the Dutch word for election ("verkiezingen") did not appear in the dataset. Although politically orientated messages were explicitly present within the dataset, where users expressed their dissatisfaction with the specific political figures of the sitting government.

Another notable distinction lies in the creative utilization of FINT by the more experienced OSINT analyst. For instance, they demonstrated creativity by performing a browser CtrlF search on raw Posts. Furthermore, the more experienced OSINT analysts displayed greater ease in integrating information from various visualisations, leading to improved responses to specific questions. One possible explanation for these differences is that thinking creatively about tool usage and merging visualisations demands more mental energy. The experienced OSINT analysts, drawing on their expertise to use heuristics and make certain assumptions regarding the Posts, can experiment and delve deeper into other aspects.

The background of the more experienced analysts might also explain the differences in scores since social media is often used for investigation. This is focused on a single detail, based on a specific task, such as finding a perpetrator, whereas this assignment was more focused on gaining insight on a more global level. Moreover, when people experience time pressure, they are more likely to fall back on their known strategies. This might also explain why the OSINT analysts used the features of FINT that were close to the visualisations they use in their current tooling.

During the exit interview, it was suggested to make use of threads, which are short conversations where users can respond to each other's Posts. Because conversations within a thread automatically form a theme, it might be easier to train a language model, since the corpora can be larger than single, unrelated Posts.

VII. Ethical Statement

The study obtained ethical approval from our local institutional review boards and ethics committees, ensuring compliance with GDPR guidelines to minimize privacy concerns. Personal data collected from participants was kept to a minimum and was anonymized (excluding optional email addresses the participants could provide to obtain a price for the best score on the assignments).

The source code of FINT, accessible via the GitHub repository [12], serves scientific purposes without any warranties. Adhering to European Data Protection Board (EDPB) guidelines, we assessed the tool's risks and impacts. Users are advised to follow similar guidelines to use FINT lawfully, transparently, and responsibly, safeguarding individuals' freedoms.

Acknowledgement

This project received partial support from the Dutch Police, who generously shared insights into their open-source intelligence work. Special appreciation is extended to the OSINT team for their invaluable contributions and insights into their operational activities.

References

[1] Isabel Ortiz, Sara Burke, Mohamed Berrada, and Hernan Saenz Cort´ es.´ *World protests: A study of key protest issues in the 21st century*. Springer Nature, 2022.

[2] Tom Postmes, Maarten van Bezouw, Susanne Tauber, and Hans van de¨ Sande. *Stampij en sociale onrust in Nederland?: Naar een beter begrip van 21e eeuws protest en oproer*. Rijksuniversiteit Groningen, 2013.

[3] Dutch National Police. Cijfers 2021: Agressie en geweld tegen politie stabiel, maar te hoog. https://www.politie.nl/nieuws/2022/april/25/cijfers-2021-agressie-en-geweld-tegen-politie-stabiel-maar-te-hoog.html, 2022. [Online; accessed 2023-02-23].

[4] Dmytro Lande, Igor Subach, and Alexander Puchkov. A system for analysis of big data from social media. *Information & Security*, 47(1):44–61, 2020.

[12] https://github.com/LMuter/FINT.git

[5] Robert David Steele. Open source intelligence. *Handbook of intelligence studies*, 42(5):129–147, 2007.

[6] Yasemin Ozkent. Social media usage to share information in communi-¨ cation journals: An analysis of social media activity and article citations. *Plos one*, 17(2):e0263725, 2022.

[7] Jyri Rajamaki and Jussi Simola.¨ How to apply privacy by design in osint and big data analytics. In *ECCWS 2019 18th European Conference on Cyber Warfare and Security*, page 364. Academic Conferences and publishing limited, 2019.

[8] Beaux Sharifi, Mark-Anthony Hutton, and Jugal K Kalita. Experiments in microblog summarization. In *2010 IEEE Second International Conference on Social Computing*, pages 49–56. IEEE, 2010.

[9] Adam Marcus, Michael S Bernstein, Osama Badar, David R Karger, Samuel Madden, and Robert C Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–236, 2011.

[10] Siqi Zhao, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. Analyzing twitter for social tv: Sentiment extraction for sports. In *Proceedings of the 2nd International Workshop on Future of Television*, volume 2, pages 11–18, 2011.

[11] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122. IEEE, 2010.

[12] Renato Kempter, Valentina Sintsova, Claudiu Musat, and Pearl Pu. Emotionwatch: Visualizing fine-grained emotions in event-related tweets. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 236–245, 2014.

[13] Bas Sijtsma, Pernilla Qvarfordt, and Francine Chen. Tweetviz: Visualizing tweets for business intelligence. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1153–1156, 2016.

[14] Laurens H.F. Muter and Remco C. Veltkamp. Analysing protest-related¨ tweets: An evaluation of techniques by the open source intelligence team. To appear in Future of Information and Communication Conference, 2024.

[15] Sydney A Gregory. Design science. In *The design method*, pages 323–330. Springer, 1966.

[16] Arun Rai. Editor's comments: Diversity of design science research. *MIS quarterly*, 41(1):iii–xviii, 2017.

[17] Ken Peffers, Tuure Tuunanen, and Bjorn Niehaves.¨ Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research, 2018.

[18] David Jones, Central Queensland University, Australia, Shirley Gregor, and The Australian National University, Australia. The anatomy of a design theory. *J. Assoc. Inf. Syst.*, 8(5):312–335, May 2007.

[19] Dan Saffer. *Designing for interaction: creating innovative applications and devices*. New Riders, 2010.

[20] Adriana Chammas, Manuela Quaresma, and Claudia Mont'Alv´ ao.˜ A closer look on the user centred design. *Procedia Manufacturing*, 3:5397–5404, 2015.

[21] Jens Rasmussen and Kim J Vicente. Coping with human errors through system design: implications for ecological interface design. *international Journal of Man-machine Studies*, 31(5):517–534, 1989.

[22] Clark Borst, John M Flach, and Joost Ellerbroek. Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human-Machine Systems*, 45(2):164–175, 2014.

[23] C Reising Dal Vernon and Penelope M Sanderson. Work domain analysis and sensors i: principles and simple example. *International Journal of Human-Computer Studies*, 56(6):569–596, 2002.

[24] Timothy R McEwen, John M Flach, and Nancy C Elder. Interfaces to medical information systems: Supporting evidenced based practice. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 335–340. IEEE, 2014.

[25] Daniel S Hall, Lawrence G Shattuck, and Kevin B Bennett. Evaluation of an ecological interface design for military command and control. *Journal of Cognitive Engineering and Decision Making*, 6(2):165–193, 2012.

[26] C Borst, M Mulder, and MM Van Paassen. Design and simulator evaluation of an ecological synthetic vision display. *Journal of guidance, control, and dynamics*, 33(5):1577–1591, 2010.

[27] Joost Ellerbroek, Koen CR Brantegem, Marinus Maria van Paassen, and Max Mulder. Design of a coplanar airborne separation display. *IEEE Transactions on Human-Machine Systems*, 43(3):277–289, 2013.

[28] Kirsten Revell, Pat Langdon, Mike Bradley, Ioannis Politis, James Brown, Simon Thompson, Lee Skrypchuk, Alexandros Mouzakitis, and Neville Stanton. Uceid-the best of both worlds: Combining ecological interface design with user centered design in a novel hf method applied to automated driving. In *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2018 International Conference on Human Factors in Transportation, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9*, pages 493–501. Springer, 2019.

[29] Javier Pastor-Galindo, Pantaleone Nespoli, Felix G´ omez M´ armol, and´ Gregorio Mart´ınez Perez. The not yet exploited goldmine of osint: Op-´ portunities, open challenges and future trends. *IEEE Access*, 8:10282–10304, 2020.

[30] Roel J Wieringa. *Design science methodology for information systems and software engineering*. Springer, 2014.

[31] Alla Kholmatova. *Design Systems: A practical guide to creating design languages for digital products*. Smashing Magazine, 2017.

[32] Will Fanguy. The benefits of shared design systems, Dec 2017.

[33] Vitaliy Ilyukha. Material ui vs bootstrap: A detailed comparison, Feb 2021.

[34] Aravind Shenoy, Anirudh Prabhu, Aravind Shenoy, and Anirudh Prabhu. Choosing lightweight frameworks for intuitive web design. *CSS Framework Alternatives: Explore Five Lightweight Alternatives to Bootstrap and Foundation with Project Examples*, pages 1–14, 2018.

[35] Jacco van der Plaat. Smid social media incident dashboard. unpublished thesis, 2021.

[36] Carmen J Ensink, Noel LW Keijsers, and Brenda E Groen. Translation¨ and validation of the system usability scale to a dutch version: D-sus. *Disability and Rehabilitation*, pages 1–6, 2022.