D6.2 FIRST ITERATION PILOTING AND VALIDATION REPORT

Deliverable No.:	D6.2
Deliverable Title:	First iteration piloting and validation report
Project Acronym:	Fandango
Project Full Title:	FAke News discovery and propagation from big Data and Artificial iNtelliGence Operations
Grant Agreement No.:	780355
Work Package No.:	6
Work Package Name:	Big data value validation through piloting
Responsible Author(s):	Lead Siren, Partic. ENG, LvT, VRT, CERTH, CIVIO, ANSA
Date:	07.02.2020
Status:	V0.9
Deliverable type:	REPORT
Distribution:	PUBLIC

REVISION HISTORY

VERSION	DATE	MODIFIED BY	COMMENTS
V0.1	07.01.2020	Jeferson Zanim (Siren/Sindice)	First draft.
V0.2	07.02.2020	Jeferson Zanim (Siren/Sindice)	Consolidated results input.
V0.5	10.02.2020	Jeferson Zanim (Siren/Sindice)	Result analysis.
V0.9	14.02.2020	Jeferson Zanim (Siren/Sindice)	Content revision and methodology update.
V1.0	21.02.2020	Jeferson Zanim (Siren/Sindice)	Partners input, methodology revision and conclusion update.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABBREVIATION	DESCRIPTION
H2020	Horizon 2020
EC	European Commission
WP	Work Package
EU	European Union

EXECUTIVE SUMMARY

This document is a deliverable of the FANDANGO project funded by the European Union's Horizon 2020 (H2020) research and innovation programme under grant agreement No 780355. It is a public report that describes the first pilot iteration of FANDANGO and its results.

FANDANGO's requirements were established in D2.3, through interview with journalists to analyze their specific needs when tackling misinformation. User stories were defined for each scenario, providing more details on the required actions and their purposes in the journalistic purposes.

Based on these user stories, D6.1 defined clear use cases and its applications in different domains, like Immigration, Climate and European Affairs. These different use cases were the main driver for the solution design and its implementation. An overview of each one is provided in Table 1.

USE CASE	DESCRIPTION OF PURPOSE
1. Verify a claim	This use case describes the process a journalist should be following when verifying whether a claim is trustworthy or rather considered as disinformation.
2. Verify a video	This use case describes the process a journalist should be following when verifying whether a video is trustworthy or rather considered as disinformation.
3. Verify an article	This use case describes the process a journalist should be following when verifying whether an article is trustworthy or rather considered as disinformation.
4. Verify an image	This use case describes the process a journalist should be following when verifying whether an image is trustworthy or rather considered as disinformation.

Table 1 - Overview of all use cases

A first version of the platform that was able to address the four different use cases was made available to users during the pilot, where the main goal was to acquire quantifiable feedback, evaluate applicability of the solution in real scenarios, assess impact of the solution in process

The main goal of this deliverable is to document and analyze the results of the first structured interaction between end users and FANDANGO's platform.

User feedback and benchmarking is essential to ensure that a software solution is effective at the proposed use cases. As such, this first pilot is focused on understanding the following: impact that FANDANGO in the day-to-day activities of journalists compared to current methods, challenges faced with the current version of FANDANGO, necessary improvements to facilitate adoption, and essential changes in the functionality.

As the project progresses, and based on the feedback provided by this iteration, further iterations shall be carried out as part of deliverable *D6.3 – Second iteration piloting and validation report*, which also expands the focus of the evaluation to technical criterias, based on the final design of the solution.

1. INTRODUCTION

FANDANGO's goal is to empower journalists when verifying different typologies of news data, media sources, social media and open data to detect publications containing misinformation and provide a more efficient and verified form of communication for European citizens.

To achieve such goal, several types of analysis have to be done by journalists to trace information source and accuracy. These can involve text, image and video analysis, requiring a number of different methods and technologies to be used in conjunction. As such, the consortium's developed a Big Data based solution, where different analysis modules enable the features provided by FANDANGO and facilitate such use cases.

To assess the effectiveness of this implementation, a pilot was planned and delivered by the consortium, engaging the user partners in real-world scenarios. The method established to execute the pilot, collect the data, analyze the results and outcomes is detailed throughout this document.

These results are fundamental inputs to the next stage of the platform and it will drive the enhancements that are required before FANDANGO platform is complete, within the scope of the project.

2. METHODOLOGY

The areas of impact to measure FANDANGO's added value compared to the current processes were defined in *D6.1 – FANDANGO Pilot execution and evaluation plan*. These are highlighted in Table 2, below.

METRIC	DESCRIPTION
1. Speed	How much faster will a user be able to perform their news verification tasks using FANDANGO versus the *as-is* situation?
2. Confidence	Journalists trust their own judgement and that of their colleagues. That is why they have high confidence in the results of their own work. This is the advantage of a manual verification process, but the disadvantage is that it's slow. FANDANGO will be very suited to address the speed parameter, but it has to guarantee a certain level of confidence in the results.
3. Number of Verifications	The average number of verifications that happen in any given newsroom, per verification, during the pilots planned in the project. Does FANDANGO increase that number? Does it manifest as longer sessions per user, or as an increased number of validations?
4. Availability	Does the sheer existence and 24/7 availability of an online news verification system like FANDANGO have any impact on the users? Manually checking a certain piece of information often requires finding human sources, people close to the subject that can verify or negate the information, calling government officials, sending emails to statistics bureaus and other time-consuming actions. When depending on other people for information, it often means being restricted by office hours, fighting for a spot in their work schedule, etc.
5. Shareability	Does FANDANGO impact the shared knowledge in the newsroom? Do the journalists share their analyzed results with colleagues? This is something we will keep track of,

	because it will help the proliferation of good news verification practices, knowledge about and traffic to the FANDANGO services.

Table 2 – Areas of Impact

Each of the use cases that were tested in the pilots were evaluated based on their relevant KPIs. The aim of the selected metrics were to provide a clear understanding of the strengths and weaknesses of FANDANGO's functionality when supporting each of the use. Therefore, the following sections define the specific KPIs tracked during the pilots to measure the impact across the defined areas.

2.1. KPIs FOR USE CASE "VERIFY A CLAIM"

AREA OF IMPACT	KPI	DESCRIPTION
Speed	Time to first impression	Time, in minutes, until the [user] has established a first impression about the possible accuracy of the claim.
Speed	Time to conclusion	Time, in minutes, until the [user] has concluded the investigation with an acceptable outcome.
Confidence	Confidence after first impression	Rate, measured from 1 to 10, where 1 means "Not confident in the outcome" and 10 means "Absolute confident of the outcome", based on the user perception after achieving a first impression.
Confidence	Confidence after conclusion	Rate, measured from 1 to 10, where 1 means "Not confident in the outcome" and 10 means "Absolute confident of the outcome", based on the user perception after reaching a conclusion for the claim verification.
Number of verifications	Number of claims verified	Total number of claims verified per day, in a given newsroom.
Number of verifications	Number of claims verified per user	Number of claims verified per newsroom user, per day.
Number of verifications	Number of tools	Number of different tools used to verify a claim.

Number of Verifications	Number of criterias checked	Number of different criteria evaluated to verify a claim, e.g. credibility of author, similar claims, validation of open data about the subject.
Availability	Waiting time	Time, in minutes, that the user has to wait on the response of other people or applications during the verification process.
Shareability	Convenience of sharing results	Convenience of making the results of a claim verification available to other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy", based on the user perception after reaching a conclusion for the claim verification.
Shareability	Convenience of accessing shared results	Convenience of accessing the results of a verification already done by other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy", based on the user perception when checking a claim that's already verified.

Table 3 - KPIs to be measured during the pilot for use case "Verify a claim"

2.2. KPIs FOR USE CASE "VERIFY A VIDEO"

AREA OF IMPACT	KPI	DESCRIPTION
Speed	Time to first impression	Time, in minutes, until the [user] has established a first impression about the possible accuracy of the video.
Speed	Time to conclusion	Time, in minutes, until the [user] has concluded the investigation with an acceptable outcome.
Confidence	Confidence after first impression	Rate, measured from 1 to 10, where 1 means "Not confident in the outcome" and 10 means "Absolute confident of the outcome", based on the user perception after achieving a first impression.
Confidence	Confidence after conclusion	Rate, measured from 1 to 10, where 1 means "Not confident in the outcome" and 10 means "Absolute

		confident of the outcome", based on the user perception after reaching a conclusion for the video verification.
Number of verifications	Number of videos verified	Total number of videos verified per day, in a given newsroom.
Number of verifications	Number of videos verified per user	Number of videos verified per newsroom user, per day.
Number of verifications	Number of tools	Number of different tools used to verify a video.
Number of verifications	Number of criterias checked	Number of different criteria evaluated to verify a video, e.g. credibility of author, timestamp.
Availability	Waiting time	Time, in minutes, that the user has to wait on the response of other people or applications during the verification process.
Shareability	Convenience of sharing results	Convenience of making the results of a video verification available to other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy", based on the user perception after reaching a conclusion for the video verification.
Shareability	Convenience of accessing shared results	Convenience of accessing the results of a verification already done by other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy", based on the user perception when checking a video that's already verified.

Table 4 - KPIs to be measured during the pilot for use case "Verify a video"

2.3. KPIs FOR USE CASE "VERIFY AN ARTICLE"

AREA OF IMPACT	KPI	DESCRIPTION

Speed	Time to first impression	Time, in minutes, until the [user] has established a first impression about the possible accuracy of the overall news article and its contents.
Speed	Time to conclusion	Time, in minutes, until the [user] has concluded the investigation with an acceptable outcome.
Confidence	Confidence after first impression	Rate, measured from 1 to 10, where 1 means “Not confident in the outcome” and 10 means “Absolute confident of the outcome”, based on the user perception after achieving a first impression.
Confidence	Confidence after conclusion	Rate, measured from 1 to 10, where 1 means “Not confident in the outcome” and 10 means “Absolute confident of the outcome”, based on the user perception after reaching a conclusion for the article verification.
Number of verifications	Number of articles verified	Total number of articles verified per day, in a given newsroom.
Number of verifications	Number of articles verified per user	Number of articles verified per newsroom user, per day.
Number of verifications	Number of tools	Number of different tools used to verify an article.
Number of verifications	Number of criteria checked	Number of different criteria evaluated to verify an article, e.g. credibility of author, similar publications, validation of claims, validation of open data about the subject.
Availability	Waiting time	Time, in minutes, that the user has to wait on the response of other people or applications during the verification process.
Shareability	Easy to share results	Convenience of making the results of an article verification available to other users, measured in a scale from 1 to 10, where 1 means “Very difficult” and 10 means “Very easy”, based on the user perception after reaching a conclusion for the article verification.

Shareability	Easy to access results	Convenience of accessing the results of a verification already done by other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy ", based on the user perception when checking an article that's already verified.

Table 5 - KPIs to be measured during the pilot for use case "Verify an article"

2.4. KPIs FOR USE CASE "VERIFY AN IMAGE"

AREA OF IMPACT	KPI	DESCRIPTION
Speed	Time to first impression	Time, in minutes, until the [user] has established a first impression about the possible accuracy of the image.
Speed	Time to conclusion	Time, in minutes, until the [user] has concluded the investigation with an acceptable outcome.
Confidence	Confidence after first impression	Rate, measured from 1 to 10, where 1 means "Not confident in the outcome" and 10 means "Absolute confident of the outcome", based on the user perception after achieving a first impression.
Confidence	Confidence after conclusion	Rate, measured from 1 to 10, where 1 means "Not confident in the outcome" and 10 means "Absolute confident of the outcome", based on the user perception after reaching a conclusion for the image verification.
Number of verifications	Number of images verified	Total number of images verified per day, in a given newsroom.
Number of verifications	Number of images verified per user	Number of images verified per newsroom user, per day.
Number of verifications	Number of tools	Number of different tools used to verify an image.
Number of verifications	Number of criterias checked	Number of different criteria evaluated to verify an image, e.g. credibility of author, timestamp.

Availability	Waiting time	Time, in minutes, that the user has to wait on the response of other people or applications during the verification process.
Shareability	Convenience of sharing results	Convenience of making the results of an image verification available to other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy", based on the user perception after reaching a conclusion for the image verification.
Shareability	Convenience of accessing shared results	Convenience of accessing the results of a verification already done by other users, measured in a scale from 1 to 10, where 1 means "Very difficult" and 10 means "Very easy", based on the user perception when checking an image that's already verified.

Table 6 - KPIs to be measured during the pilot for use case "Verify an image"

These KPIs were assessed during the pilot, which was executed by ANSA, CIVIO and VRT. Each partner provided at least three different people and a total of 64 hours of effort, broken down into the 4 different use cases, as planned by the consortium and shown in Figure 1**Errore. L'origine riferimento non è stata trovata.**.

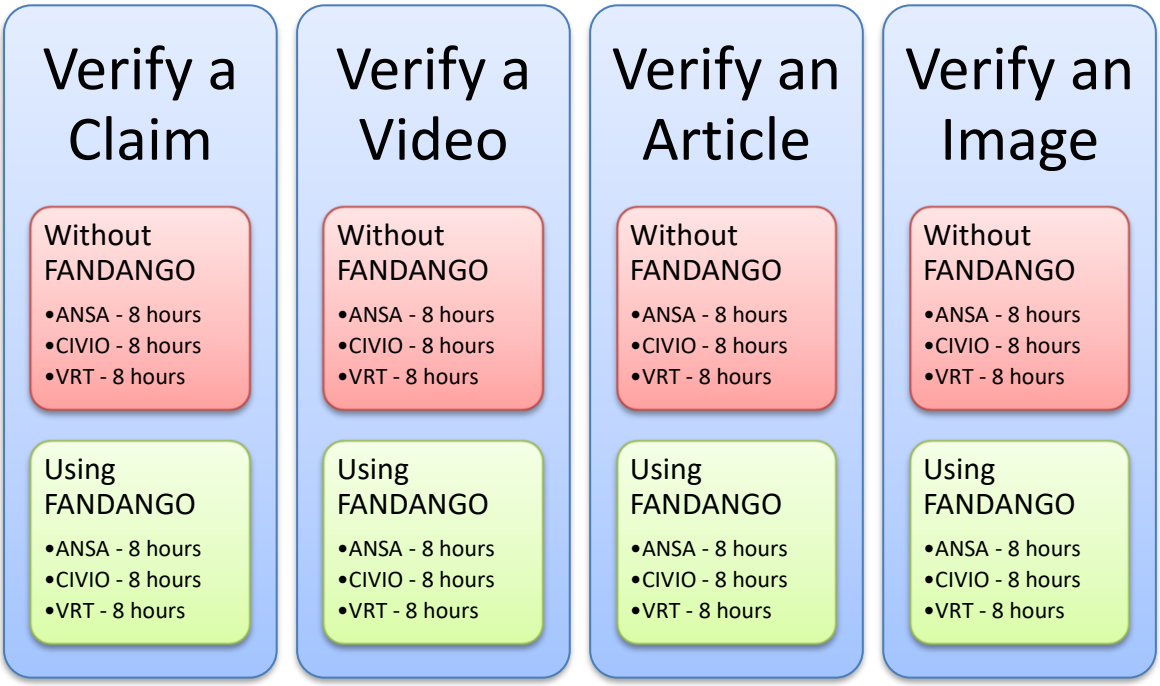

Figure 1 - Pilot Effort per Use Case

In order to measure the specific impact of FANDANGO, as outline in D6.1, the pilot was executed in two phases.

In phase 1, users were benchmarked while performing each use case without the use of FANDANGO platform, as it is currently performed.

Later, on phase 2, the same KPIs were measured with users adding FANDANGO to their process. While the use of FANDANGO was mandatory, it was not exclusive, which means that other tools could also be used. That allowed us to measure if FANDANGO is enough to provide enough confidence in the analysis or it is more often used to complement the verification process done in other tools.

3. CHALLENGES

The delivery of the first version of FANDANGO platform with support to all 4 use cases was delayed in relation to the original planning, which required the first pilot to start later than anticipated. No impact on further deadlines is expected as consequence of this delay as other activities could progress regardless of his minor delay.

While FANDANGO is designed to be an intuitive tool, it still required out user partner to be trained to use its functionality during the pilots. They then retrained the participants of the pilot to be proficient in the usage of the platform. While training was simplistic and mostly possible to be done in a self-guided manner through the written material, it would increase adoption if usage was self-explanatory, at least in its core functions.

Another imporant point highlighted by the participants was that the Advanced Analytics was too complicated, even with the limited training material and didn't provide the value they were expecting during the analysis.

During the pilot execution, there was an initial difference of understanding between the user partners regarding the method to measure the KPIs during the pilot. That produced inconsistent results initially, but it was discussed during one of our weekly meetings focused on the pilot execution, and the necesseary activities were redone by the partners to ensure consistency of the information collected.

While the information produced during the pilot is extremely relevant to drive the next steps in its implementation and validate the resources produced so far, the volume of data collected is not sufficient to provide statistically valid results that would assert the effectiviness of the platform.

4. RESULTS

As an outcome of the first pilot, FANDANGO measured the defined KPIs on each execution of the use cases. The complete results as well as the full analysis is available in the spreadsheet below.

Pilot Execution Markings V1_3.xlsx

Additionally, user partners collected feedback from the users regarding usefulness of the solution based on their news verification process and compared it with existing FANDANGO functionality. Based on the most checked criteria during the pilot, results can be seen in Figure 2.

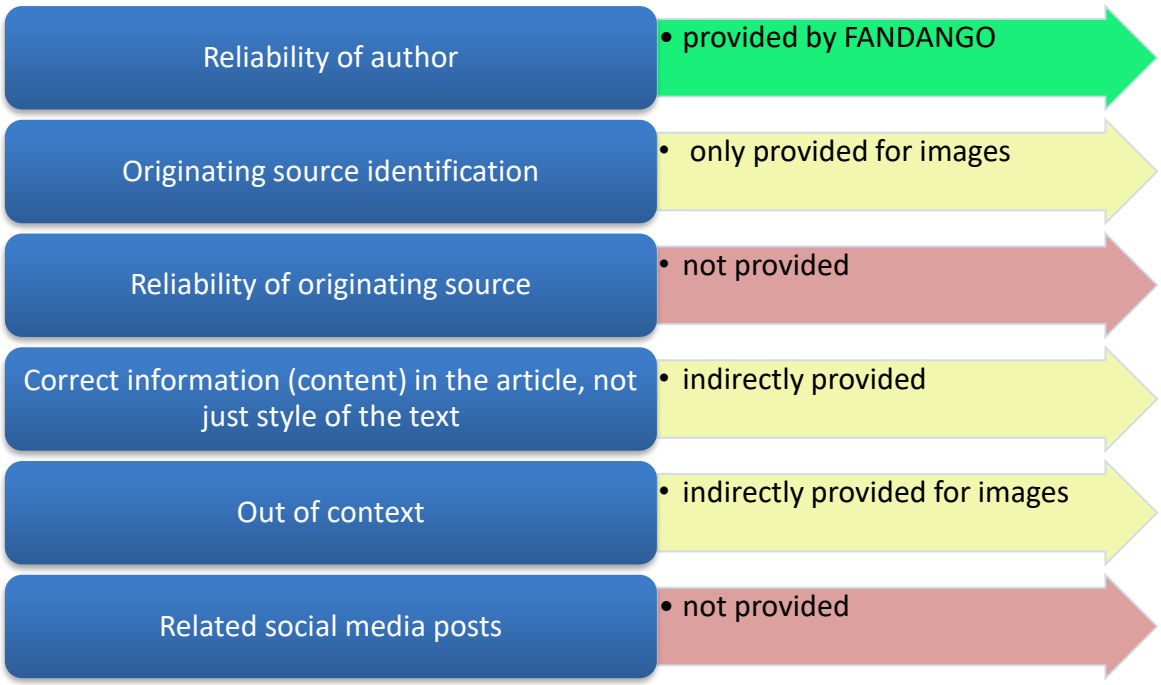

Figure 2 - Most checked criteria

Authors are scored by the system based on network analysis considering previous publications, publishing affiliations and collaboration with other authors in order to provide a trustworthiness score. Since this is one of the most checked criteria by journalists, such functionality has direct impact in their process, improving confidence and reducing effort.

Since content is often re-used in different publications, an important criterion during the evaluation is the identification of the original source of published material. FANDANGO helps journalists identifying original images through FANDANGO's TNT module, which uses reverse image search techniques to find previous occurrences of similar images and build a comprehensive timeline, providing users with contextual information regarding possible sources of the image. This type of analysis is currently limited to images though and does not include other types of media that are also analyzed in the investigation process.

Another step in the assessment process is the analysis of reliability of the original source. Once the origin has been identified for certain medias or claims, it is important to assess its credibility. From the end-user perspective, this was a type of analysis that was not yet provided by the system during the pilot. However, since FANDANGO provides author and publisher scoring, it is possible to analyze the source credibility by starting an analysis of the original publication. Some clarification in this process might be required, as well as adjust to the UI to improve this behavior in further pilots.

Analyzing the accuracy of the information provided in publications is a significant part of the verification. This is referred to as "claims", which needs to be verified. Since there are multiple organizations around the globe that focus on reviewing claims, and this process in itself isn't part of FANDANGO's scope, the current version of the platform has a Claim Search functionality, which allows users to input a specific claim to find if a previous review has been done. This feature helps journalists benefit from the work already being done by other groups to speed up the verification process.

Through the same way process that FANDANGO assists users to find the original source of an image, it also enables them to identify if an image is being used out of context by back tracing the original scenario where it was used. While it has a significant impact in this type of analysis, it is also limited to only images, offering partial coverage for the different types of medias analyzed.

Lastly, journalists also look for publications in social medias to check the credibility of certain contents and sources. However, due to the challenges of legally processing social media content, both regarding the scale of such data and the strict terms of services enforced by the controlling corporations, the automatic analysis of such content is outside of the scope of FANDANGO's deliverables. However, FANDANGO does offer certain features that can facilitate the analysis of social media posts, like the claim review search and the image timeline.

Overall, this feedback highlights that FANDANGO has an impact on most parts of the verification process. Nonetheless, it can still be further improved to assist journalists in their mission.

4.1. VERIFY A CLAIM

In this use case, journalists are looking to validate the content of an specific claim. These could be originated from a publication, a social media post, a video or a statement made on a public speech. FANDANGO provides a search tool that allows users to search for reviewed claims published by multiple fact-check organizations.

In Figure 3, we can see the impact of such functionality at reducing the analysis time. The use of FANDANGO platform reduced to less than half the original time to reach a first impression and a conclusion about a certain claim.

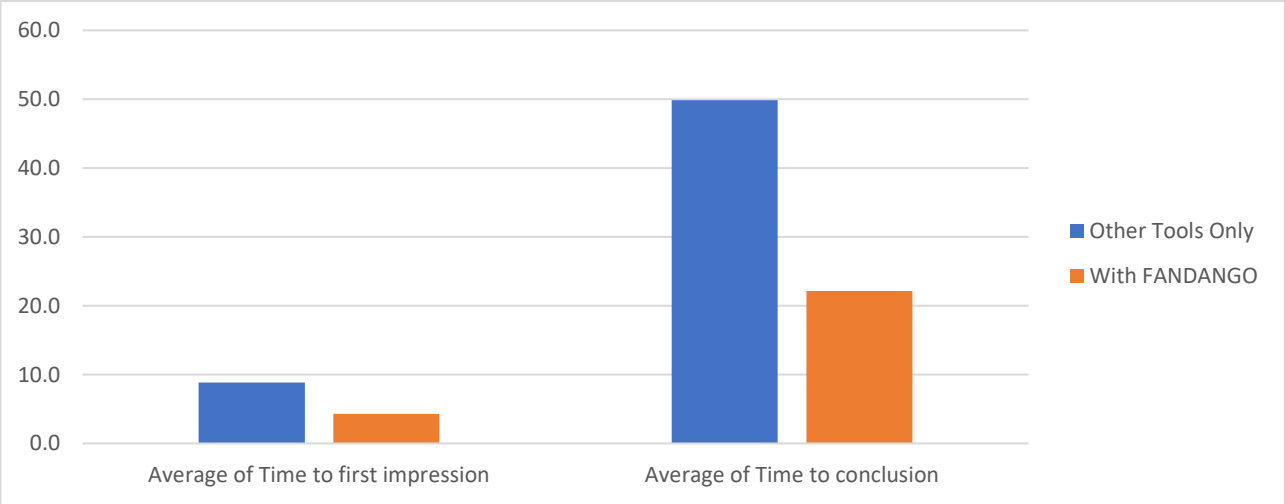

Figure 3 - Verify a Claim: Speed

However, when it comes to the confidence of the user in the analysis, the results don't show significant impact and the variances are within the expected margin of error, as we can observe in Figure 4.

Figure 4 - Verify a Claim: Confidence

Figure 5 highlights that FANDANGO acts as an additional tool when analyzing claims, instead of replacing others. Nonetheless, when being used, it requires less criteria to be analyzed before a conclusion and reduces waiting times, normally caused by having to call or email third parties to search for information.

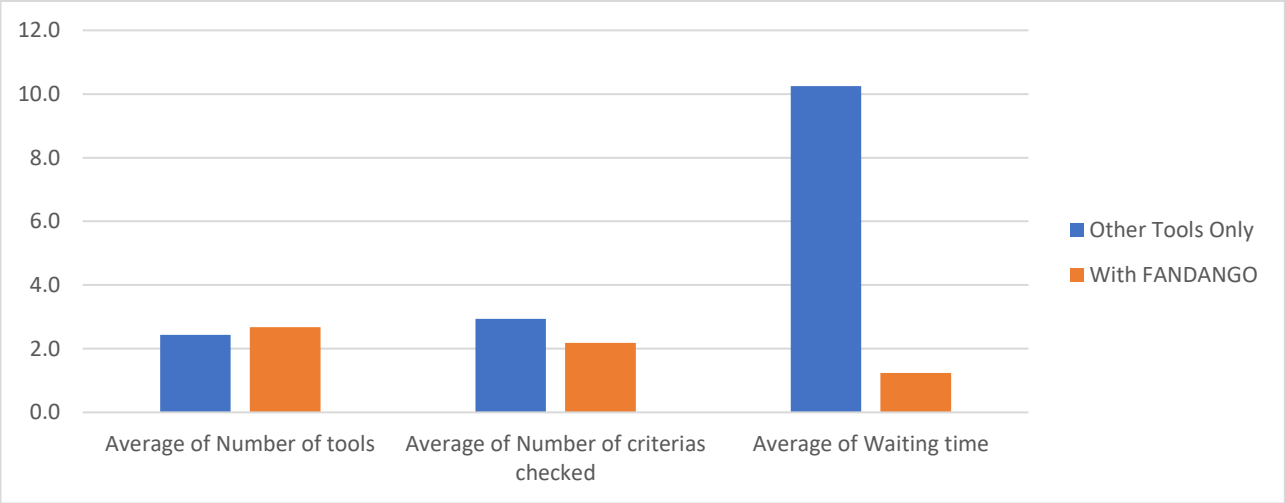

Figure 5 - Verify a Claim: Verifications and Waiting Time

These results are quite promising for the claim verification process, as FANDANGO seems to already be effective support tool for this process.

Partner Specific Results

When assessing the impact on reducing analysis time per partner, in Figure 6, we can observe that despite the difference in analysis time across the different organizations, the impact of FANDANGO has been consistent across all three.

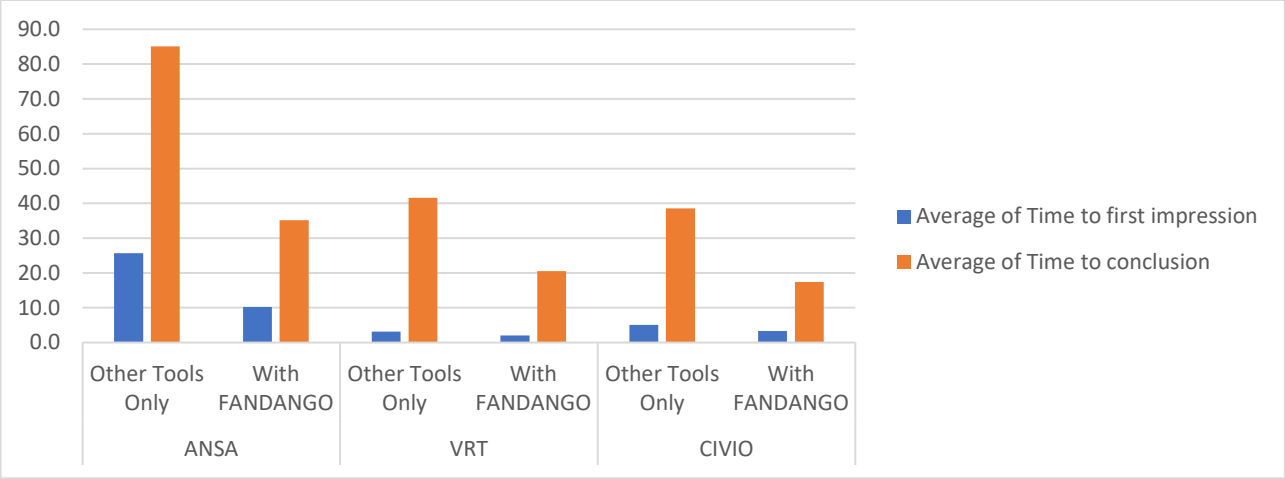

Figure 6 - Verify a Claim per Partner: Speed

When observing each organization individually, the impact in confidence after the analysis also seems to not have been significant, which is in-line with the overall result. This can be seen in Figure 7.

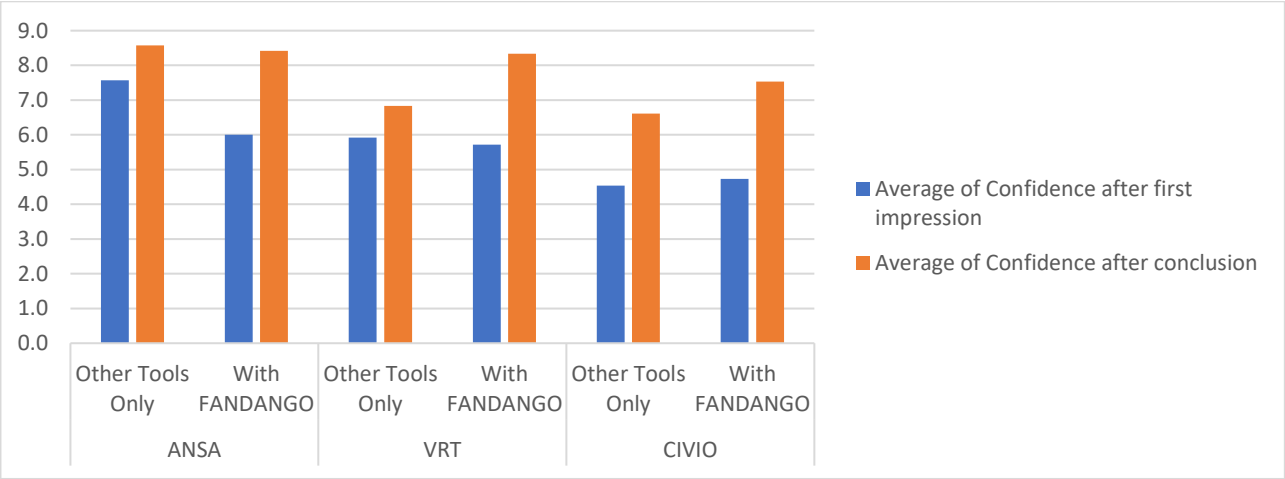

Figure 7 - Verify a Claim per Partner: Confidence

When it comes to the overall waiting time shown in Figure 8, we can notice that there was no impact for CIVIO as they don't have waiting time in their overall process, therefore, not leveraging this aspect. ANSA and VRT significantly reduce their waiting time though.

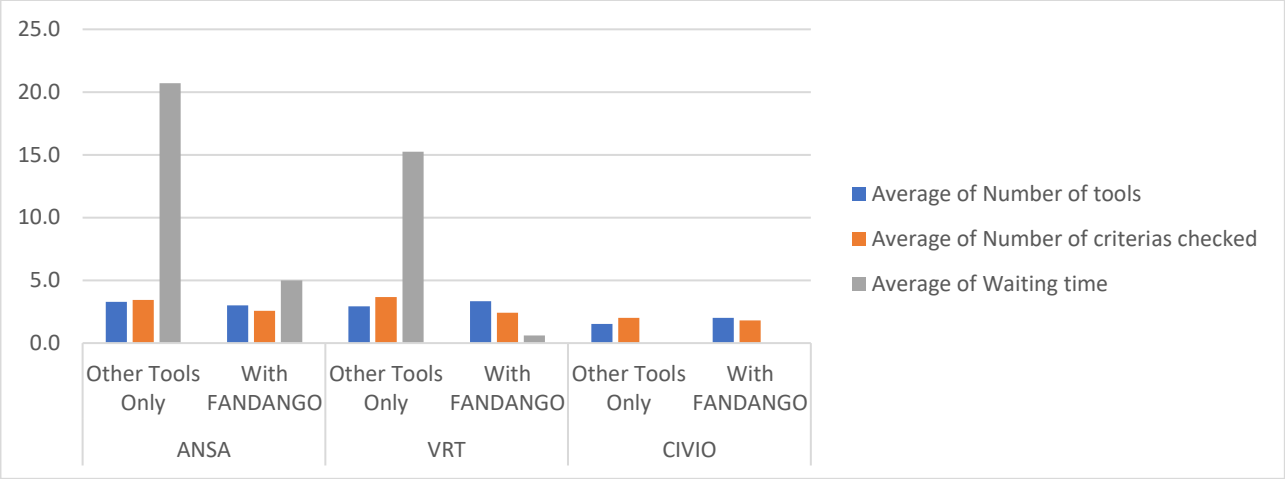

Figure 8 - Verify a Claim per Partner: Verifications and Waiting Time

CORRELATION ANALYSIS

Across the different criteria evaluated, most of the correlations identified are expected, like a longer time to conclusion when the waiting time is also longer, or a larger number of tools being related to a higher number of criteria checked.

However, one notable correlation is the fact that users seems to find it more convenient to share results when they are more confident in the results of analysis. While this behavior seems to be unrelated to the usage of FANDANGO, improving user confidence might have a significant impact on having analysis results published. As such if FANDANGO improves its ability to increase user confidence, it might also contribute to journalists sharing the information being produced.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1- Time to first impression	1.000									
2- Time to conclusion	0.703	1.000								
3- Analysis Outcome	-0.135	-0.072	1.000							
4- Confidence after first impression	0.230	0.135	-0.214	1.000						
5- Confidence after conclusion	0.100	-0.148	-0.169	0.465	1.000					
6- Number of tools	0.173	0.378	-0.048	0.115	0.243	1.000				
7- Number of criterias checked	0.243	0.378	-0.074	0.194	0.117	0.541	1.000			
8- Waiting time	0.410	0.667	0.011	0.148	-0.104	0.495	0.282	1.000		
9- Convenience of sharing results	0.357	0.001	-0.304	0.383	0.730	0.152	0.161	-0.090	1.000	
10- Convenience of accessing shared results	0.400	0.007	-0.245	0.359	0.606	0.031	0.028	-0.105	0.853	1.000

Table 7 - Verify a Claim KPI Correlation

4.2. VERIFY A VIDEO

When verifying a video, journalists are looking for two main factors: authenticity of the video and original context of the video. FANDANGO's functionality is currently focused exclusively in the first one, not yet providing context analysis.

The video analysis tool offers a series of analysis to assist users verifying if a video has been tampered with.

In certain cases users reported having certain difficulties interpreting the results of the analysis, despite a brief training in the functionality. Therefore, while the verifications to do provide state-of-the-art analysis, it might still need usability improvements to be more effective at aiding end-users in this process.

In Figure 9 we can notice that FANDANGO only marginaly contributed to accelerating the analysis of videos, which usually take substantial time. Yet, it is an initially positive result that needs to be validated with a large data sample to account for the margin of error.

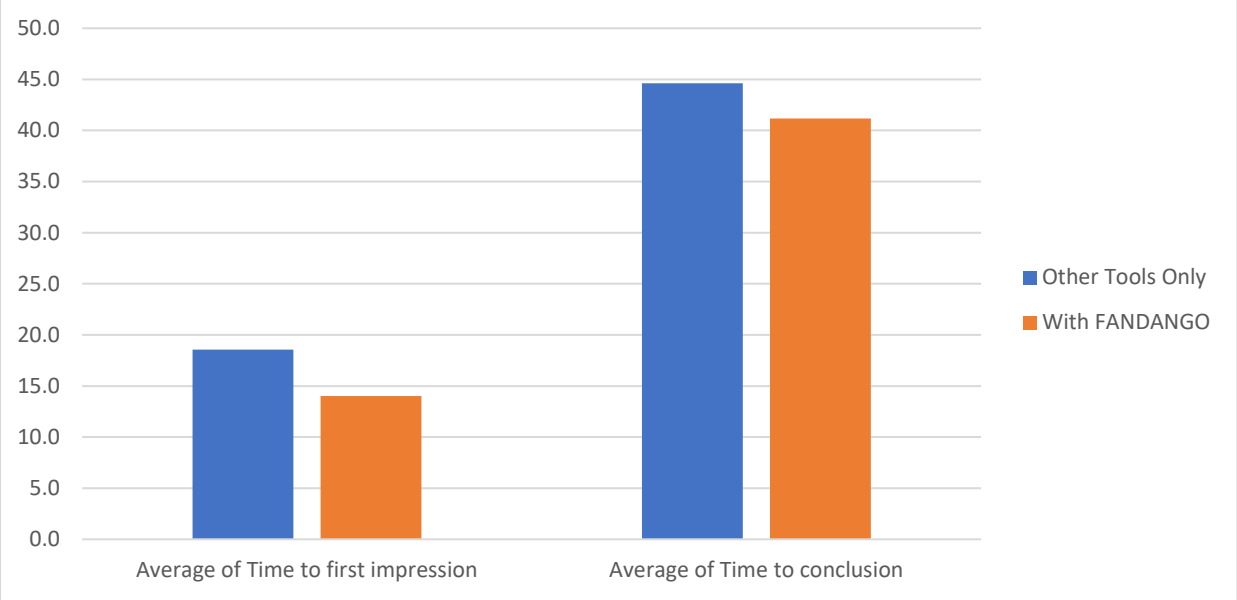

Figure 9 - Verify a Video: Speed

As in the Speed area, the data also shows only a negligible impact in confidence, shown in Figure 10.

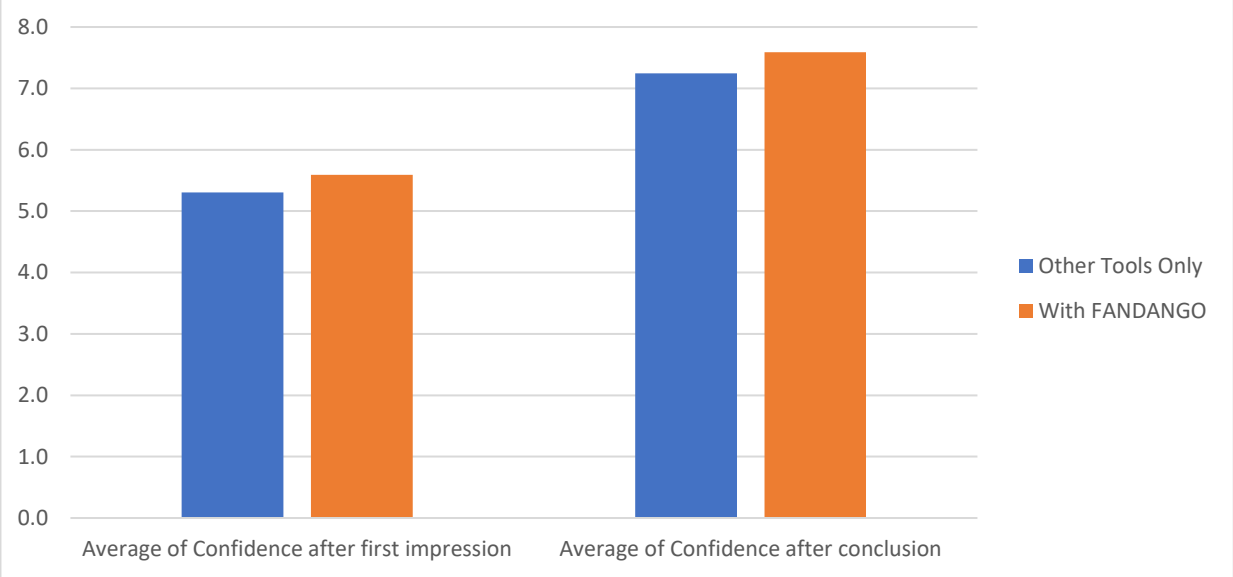

Figure 10 - Verify a Video: Confidence

Perhaps the most interesting part of the video analysis results is seen in Figure 11, where it possible to observe that despite the time to conclusion shown before being constant, when using FANDANGO the waiting time was much longer. Therefore, it is likely that FANDANGO reduced the required effort to analyze a video, but due to the time it takes for a video to be processed by the platform, the duration of the activity

ended up being similar. This opens the potential to reduce the overall analysis time by improving processing performance and descreasing waiting time.

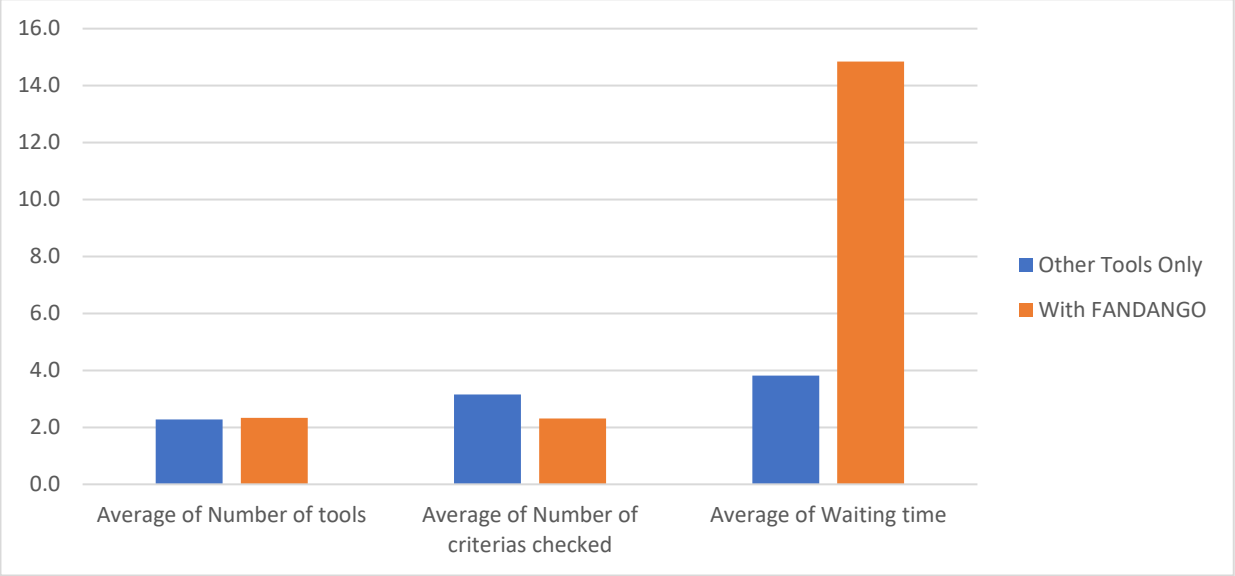

Figure 11 - Verify a Video: Verifications and Waiting Time

Results indicate that with performance improvements and making the analysis results easier to understand, the video analysis tool would significantly impact the overall process.

Partner Specific Results

When broken down by organization in Figure 12, it is possible to notice that overall process and the impact of FANDANGO are drastically different across them. Overall, ANSA seems to take longer analyzing videos and, therefore, had a time reduction using FANDANGO. On the other side of the spectrum, VRT saw an increase, which might be associated to the processing time it takes for a video to be analyzed.

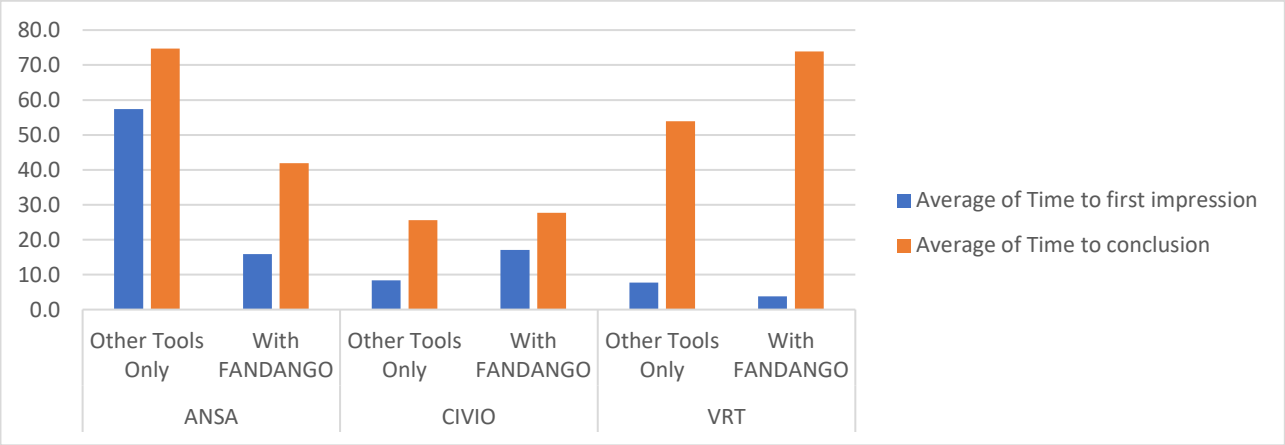

Figure 12 - Verify a Video per Partner: Speed

Regarding confidence, in Figure 13, impact is consistent across different partners, despite their overall results variying.

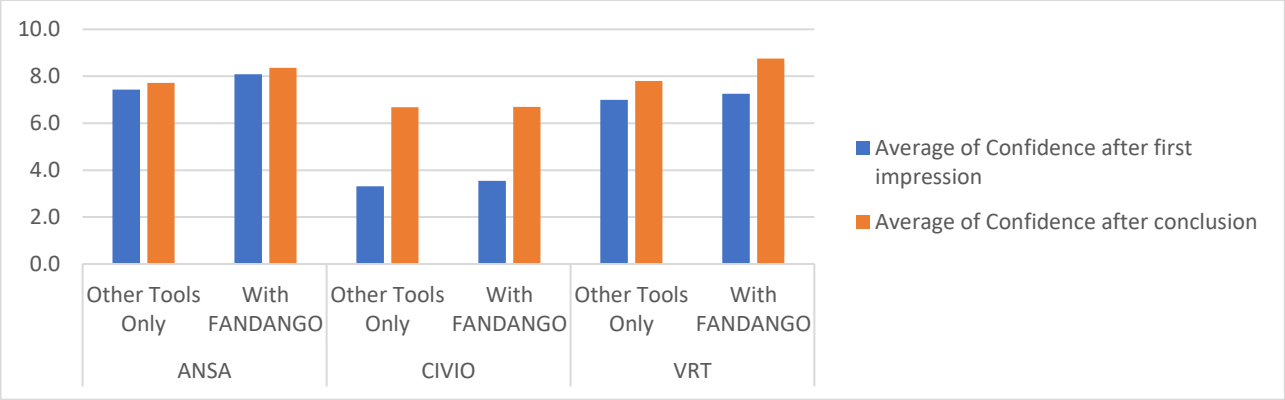

Figure 13 - Verify a Video per Partner: Confidence

Consistent with the Speed KPIs, we can see on Figure 14 that the impact on average waiting time was positive for ANSA, but negative for CIVIO and VRT.

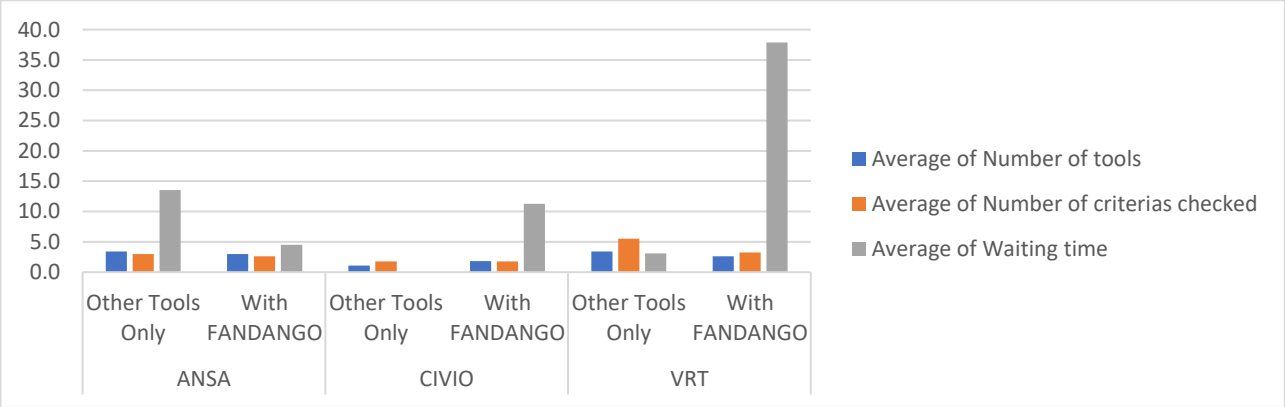

Figure 14 - Verify a Video per Partner: Verifications and Waiting Time

Further investigation is required to understand the causes of time discrepancies across the different partners. It could be related to their processes or to the difference in videos chosen for the analysis. A more strict criteria to sample the videos analyzed during the pilot might need to be defined for the next iteration.

CORRELATION ANALYSIS

No significant correlation has been observed between the different criteria, other than obvious factors like a longer waiting time often correlating to longer time to conclusion.

	1	2	3	4	5	6	7	8	9	10
1- Time to first impression	1.000									
2- Time to conclusion	0.560	1.000								
3- Analysis Outcome	-0.206	-0.106	1.000							
4- Confidence after first impression	0.069	0.358	-0.032	1.000						
5- Confidence after conclusion	-0.129	0.263	0.204	0.557	1.000					
6- Number of tools	0.310	0.652	-0.011	0.372	0.432	1.000				
7- Number of criterias checked	0.079	0.530	0.089	0.169	0.360	0.760	1.000			
8- Waiting time	0.246	0.654	0.039	0.194	0.204	0.206	0.111	1.000		
9- Convenience of sharing results	0.223	0.253	-0.066	0.560	0.589	0.419	0.213	0.046	1.000	
10- Convenience of accessing shared results	0.238	0.195	-0.301	0.517	0.277	0.194	-0.089	0.097	0.686	1.000

Table 8 - Verify a Video KPI Correlation

4.3. VERIFY AN ARTICLE

Article verification, which involves the full analysis of a certain publication, is provided by FANDANGO platform through a combination of different analysis, involving the same validations of all other use cases combined: claims, authors, publishers, images and videos.

As such, this use case is directly impacted by the performance of the other use cases as well as its own uniques characteristics. Overall, the results for the article analysis are very similar to the claims use case.

In Figure 15 we can observe that FANDANGO has a significant impact in reducing the required time to analyze articles.

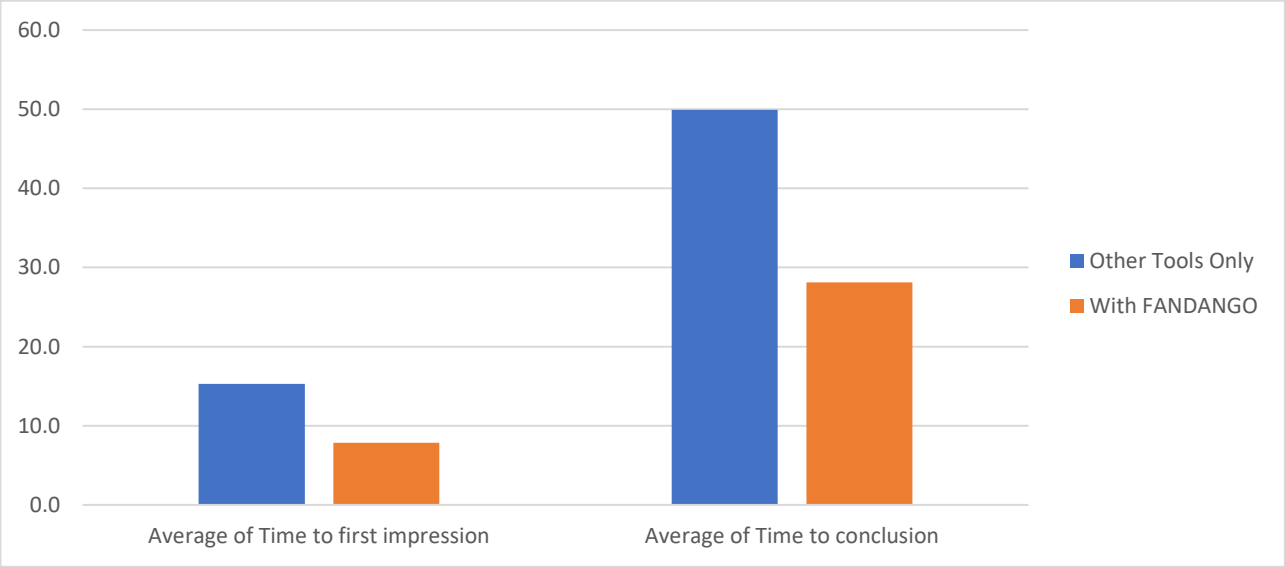

Figure 15 - Verify an Article: Speed

However, similarly to the claims use case, there has been no noticeable impact in the confidence at the end of the analysis, as shown in Figure 16. This could be related to the feedback provided by users regarding the difficulty in understanding how FANDANGO scores are produced. It is their understanding that despite a score being attributed to the linguistic profile of the publication, the authors and the publisher, without the exact criteria used to produce such scores, they can't rely on that information to make a decision.

Figure 16 - Verify an Article: Confidence

Also similarly to the claims use case, when it comes to the article verification, shown in Figure 17, FANDANGO acts as a support tool instead of a replacement. Nonetheless, it severely reduces waiting times in this process.

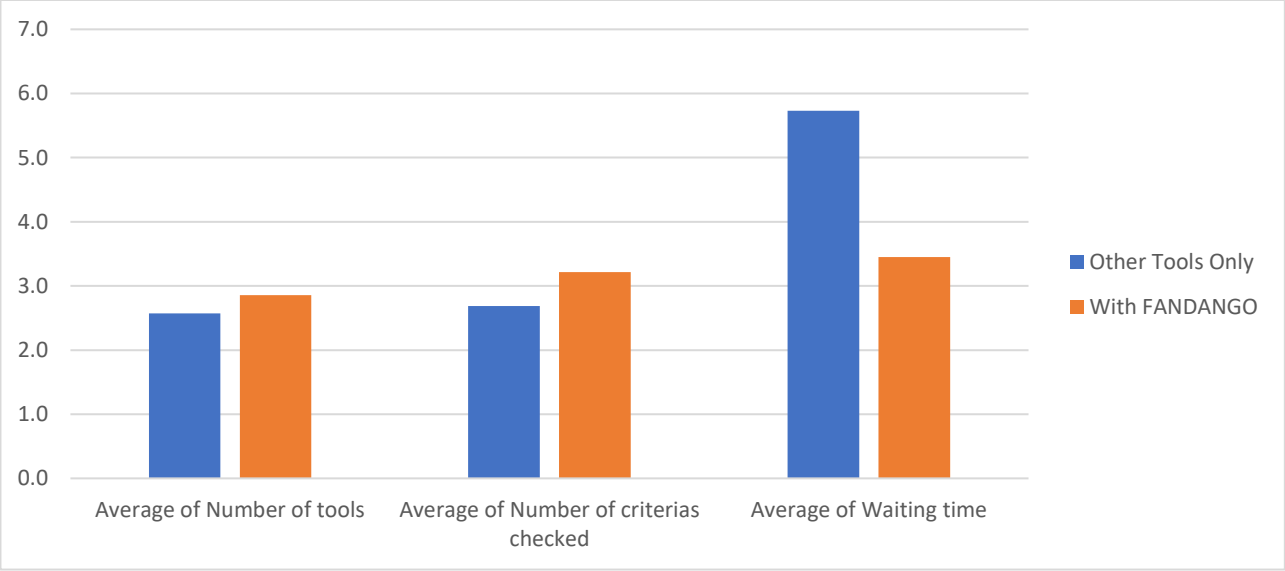

Figure 17 - Verify an Article: Verifications and Waiting Time

PARTNER SPECIFIC RESULTS

Observed in Figure 18, positive impact of FANDANGO is more prominent on ANSA, but it can also be noticed for CIVIO. VRT though, shows only a negligible variance in its results.

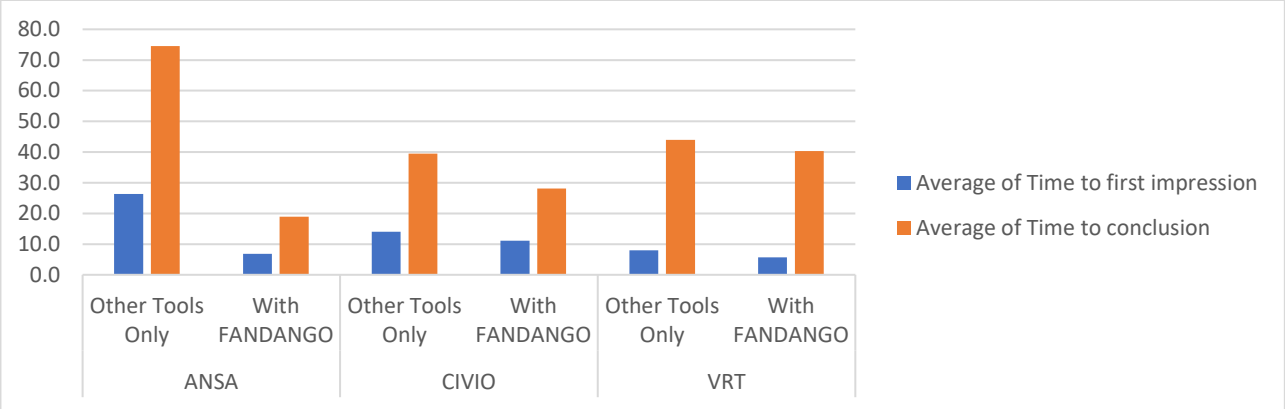

Figure 18 - Verify an Article per Partner: Speed

Results regarding confidence per partner show very small variances, either positive or negative depending on the organization. Overall they were not significant in one way or another according to Figure 19.

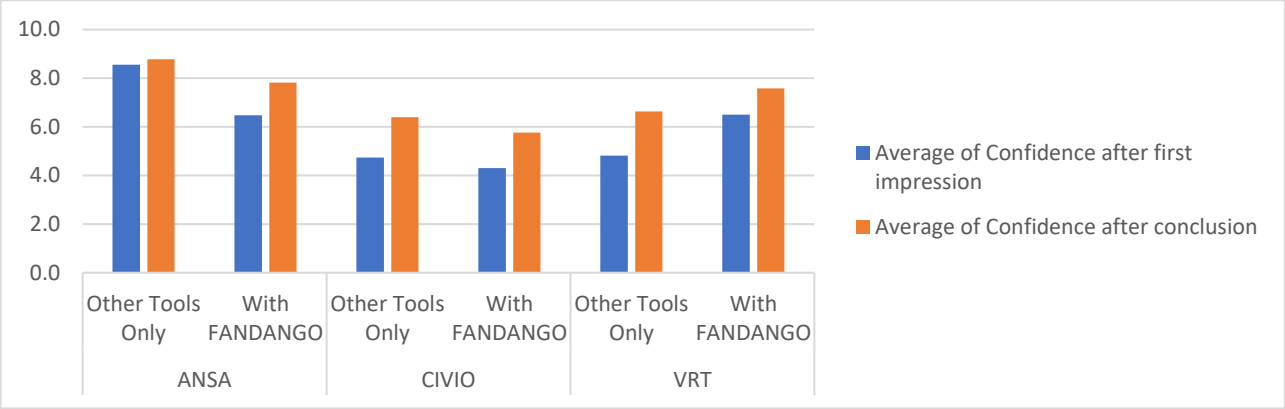

Figure 19 - Verify an Article per Partner: Confidence

On Figure 20, we can notice that the impact on waiting time is completely different across all three organizations, with ANSA observing a significant reduction, CIVIO having no impact and VRT showing an increased in waiting time. Further investigation is required to identify the main causes in this discrepancy.

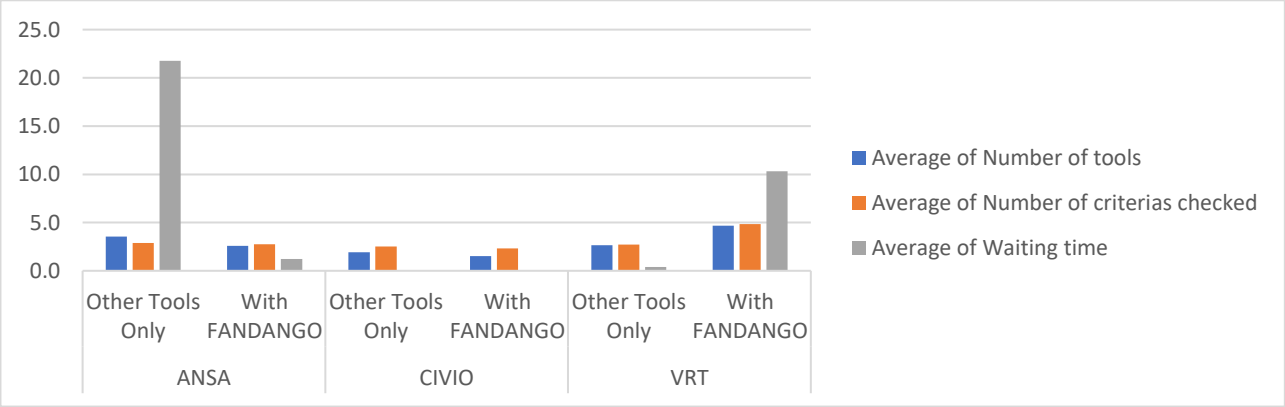

Figure 20 - Verify an Article per Partner: Verifications and Waiting Time

As in the claims verification use case, the article analysis also seems to grealy benefit from using FANDANGO. Improvements still need to be made on the scoring aspects of the solution to increase user confidence, but initial results are already promising.

CORRELATION ANALYSIS

Once again, the only significant and non-expected correlation is the increased confidence boosting the easiness of sharing results. As in the claim use case, this can be further explored to improve the analysis process by fostering collaboration.

	1	2	3	4	5	6	7	8	9	10
1- Time to first impression	1.000									
2- Time to conclusion	0.679	1.000								
3- Analysis Outcome	-0.228	-0.160	1.000							
4- Confidence after first impression	0.146	0.176	0.074	1.000						
5- Confidence after conclusion	0.059	-0.105	0.176	0.653	1.000					
6- Number of tools	0.002	0.242	0.205	0.249	0.159	1.000				
7- Number of criterias checked	-0.075	0.218	0.299	0.093	0.127	0.739	1.000			
8- Waiting time	0.525	0.762	-0.073	0.327	0.002	0.212	0.302	1.000		
9- Easy to share results	0.068	-0.059	-0.036	0.598	0.615	0.196	0.077	0.143	1.000	
10- Easy to access results	0.053	-0.059	0.058	0.619	0.566	0.209	0.114	0.203	0.910	1.000

Table 9 - Verify an Article KPI Correlation

4.4. VERIFY AN IMAGE

Image verification is done by users as part of a publication or individually, when it is used without complementary text. FANDANGO provides tools to analyze the authenticity of an image through machine learning processing, as well as original source identification through the timeline. Indirectly, users can also use the timeline to verify if the image is being used in the appropriate context.

In Figure 21 it is possible to observe that FANDANGO has a marginal impact at reducing analysis time for the images. While results are positive, it isn't as significant when considering the margin of error for the sample analyzed.

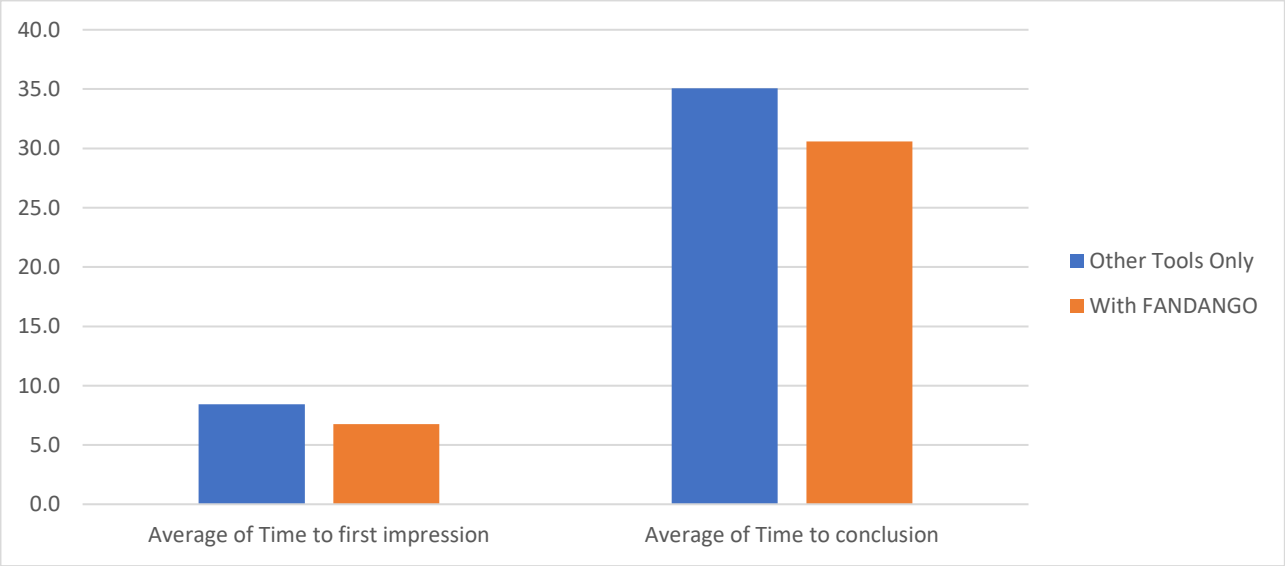

Figure 21 - Verify an Image: Speed

The user confidence was unaffected by the usage of FANDANGO, as shown in Figure 22.

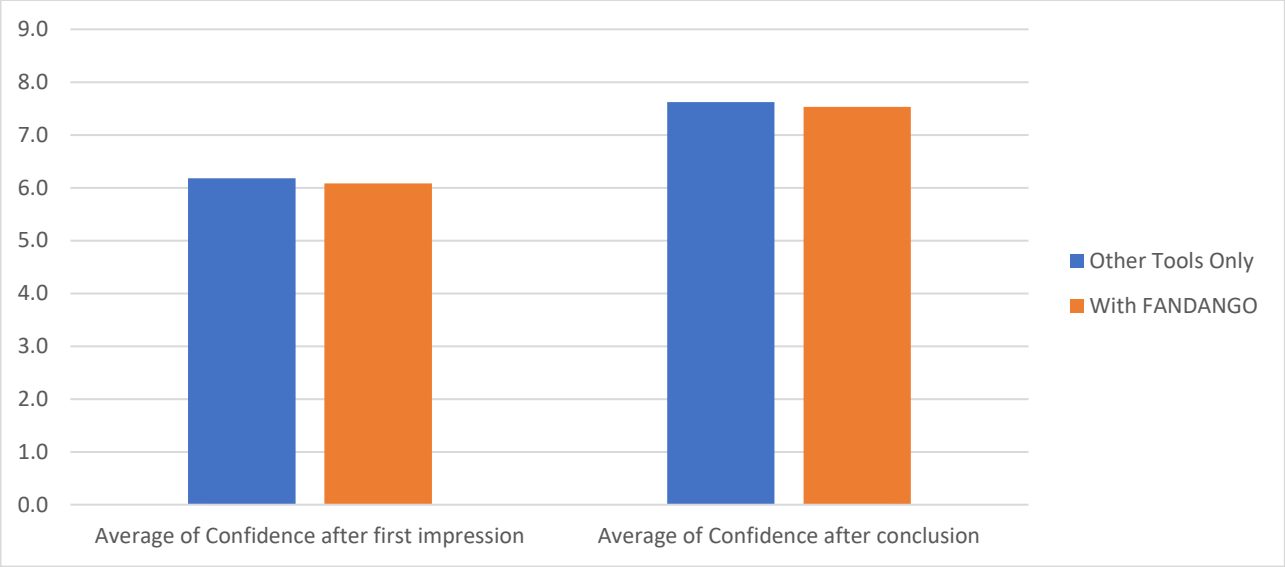

Figure 22 - Verify an Image: Confidence

Similarly to the video use case, using FANDANGO for images seems to have reduced the overall effort of the analys as the conclusion time was reduced despite the significant increased in waiting time shown in Figure 23.

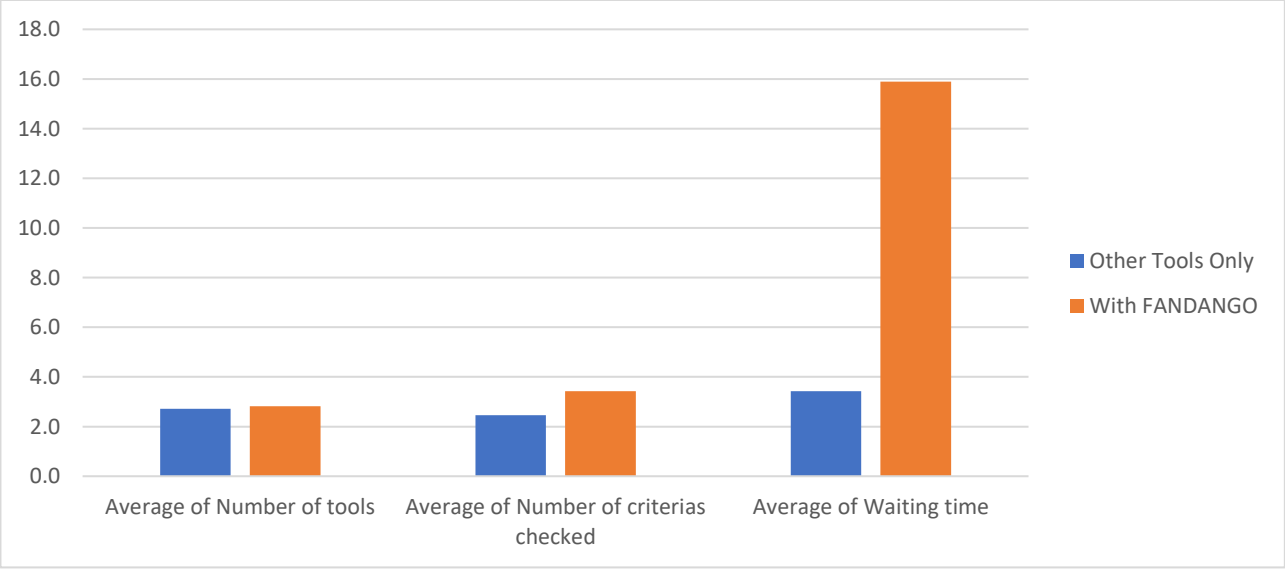

Figure 23 - Verify an Image: Verifications and Waiting Time

The image verification functionality in FANDANGO seems promising, but might still require improvements to boost user confidence and reduce waiting times. Nonetheless, it seems to already reduce required effort overall.

Partner Specific Results

Individual results per partners did vary during the pilot, as shown in Figure 24. ANSA greatly benefited from the reduction in analysis time, while CIVIO had no impact and VRT saw an overall increase.

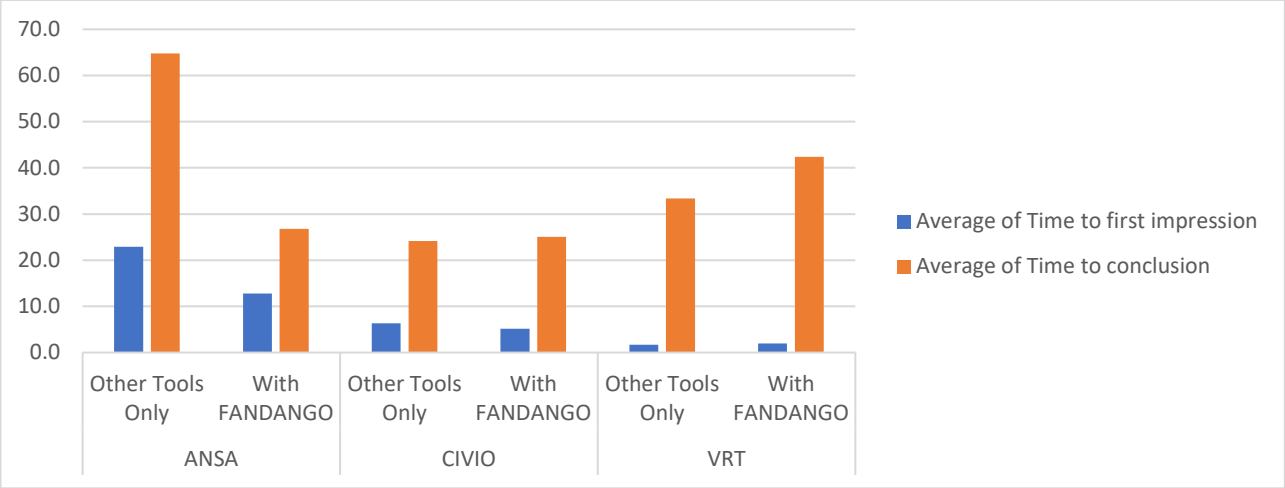

Figure 24 - Verify an Image per Partner: Speed

Based on the individual results shown in Figure 25, VRT seems to have had a small decrease in confidence when using FANDANGO to analyze images. Some of the users did report that specific types of obvious tampering with the images were not being flagged by the system, which might have reduced their confidence in the results. Samples were provided to technical partners to verify possible causes.

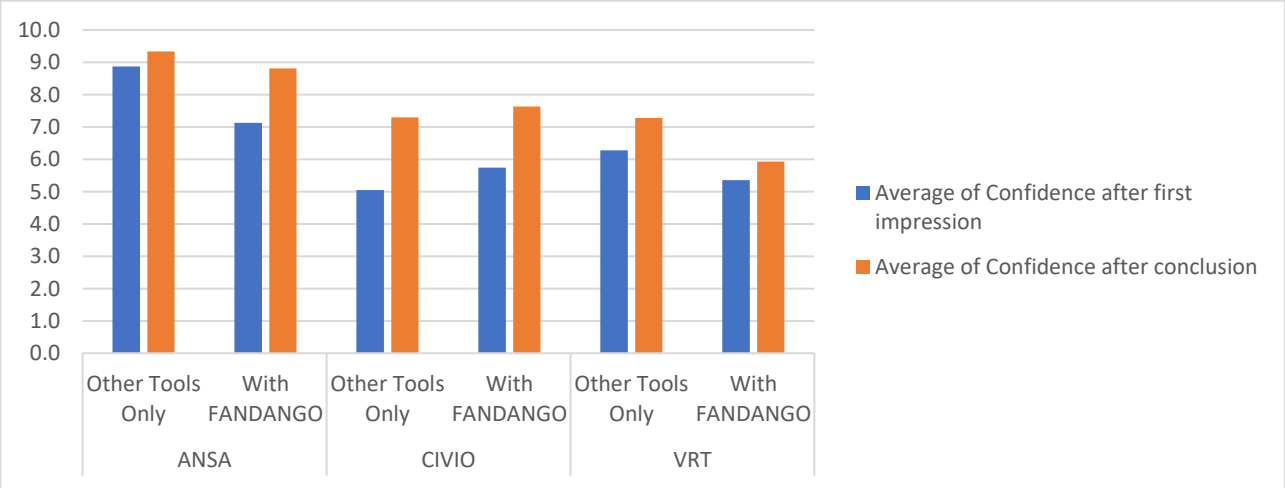

Figure 25 - Verify an Image per Partner: Confidence

Results were also inconsistent across different partners regarding waiting times. These might be related to the samples used or the methodology of each organization. In any case, this will be further investigated.

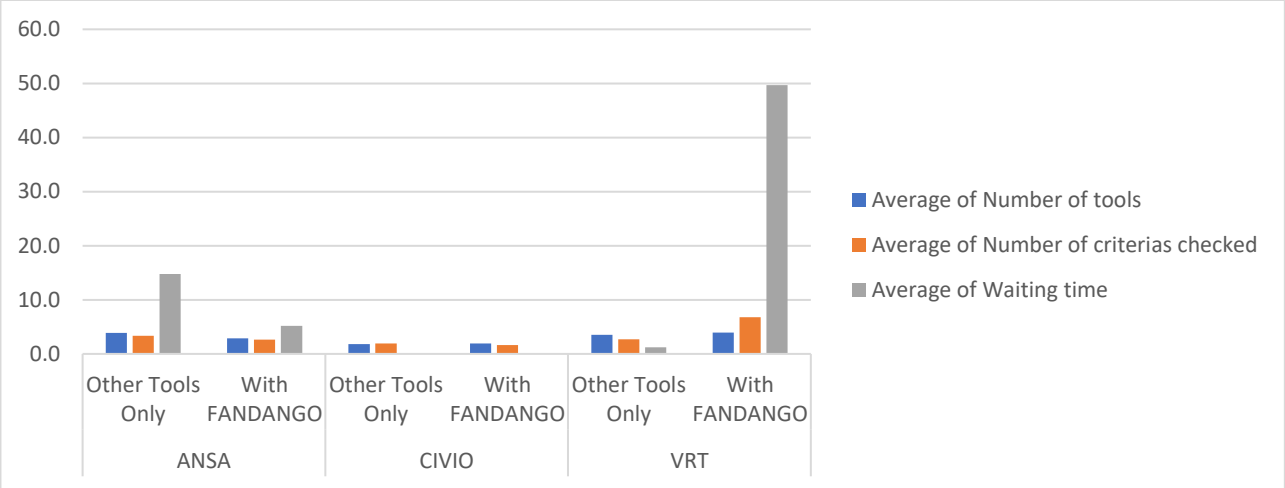

Figure 26 - Verify an Image per Partner: Verifications and Waiting Time

CORRELATION ANALYSIS

No significant non-expected correlations can be observed in the results of the image analysis use case.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1- Time to first impression	1.000									
2- Time to conclusion	0.396	1.000								
3- Analysis Outcome	-0.342	-0.125	1.000							
4- Confidence after first impression	0.130	-0.183	-0.254	1.000						
5- Confidence after conclusion	0.226	-0.414	-0.209	0.721	1.000					
6- Number of tools	0.101	0.613	0.176	0.016	-0.258	1.000				
7- Number of criterias checked	-0.074	0.188	0.067	-0.047	-0.236	0.403	1.000			
8- Waiting time	-0.027	0.267	-0.087	-0.004	-0.216	0.214	0.039	1.000		
9- Convenience of sharing results	0.446	0.028	-0.070	0.517	0.623	0.186	-0.048	0.011	1.000	
10- Convenience of accessing shared results	0.516	0.163	-0.045	0.430	0.380	0.245	0.125	0.050	0.801	1.000

Table 10 - Verify an Image KPI Correlation

5. CONCLUSION

In this deliverable we presented the results of the first FANDANGO pilot, highlighting feedback from the users as well as benchmarks collected during the execution. Given these results, the 4 priorities for the users regarding the biggest positive impact in the current new verification process would be as shown in Figure 27.

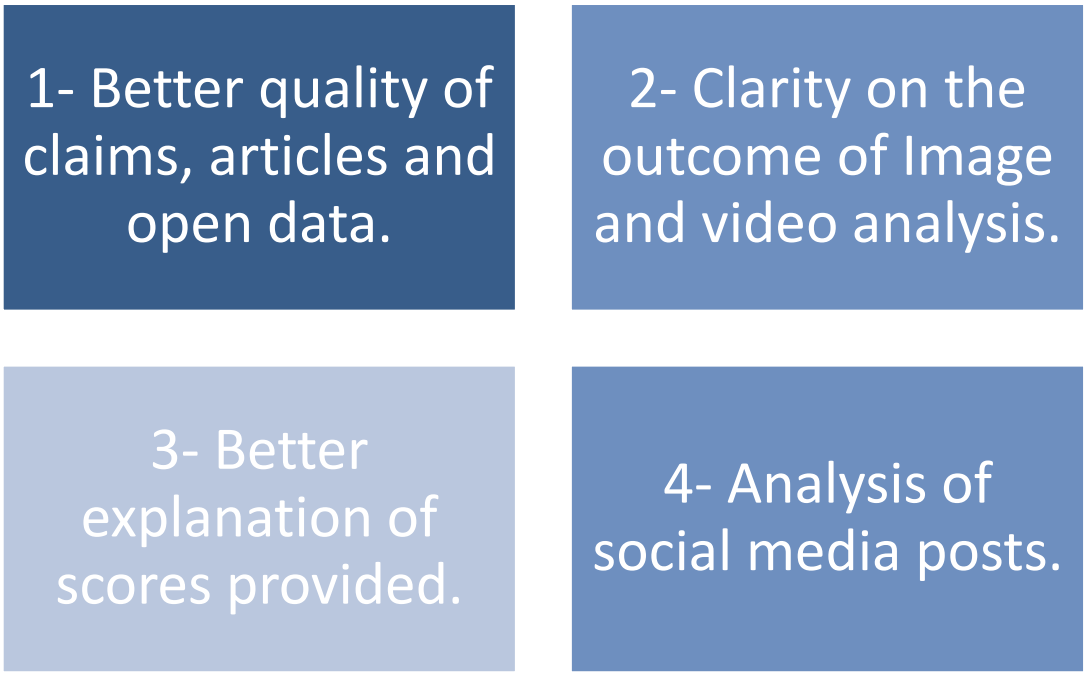

Figure 27 - Users Priorities

1. Users reported that articles, claims and open data suggested most of the time weren't relevant to their analysis. This was mostly caused by the data available being limited to a certain period in time that didn't always coincided with the period the user was analyzing.
2. Image and Video analysis results were too complex for users to understand and need clarification to make it simpler to use.
3. When analyzing an article, users want to understand the factors utilized to calculate the scores of text linguistic characteristics, authors and publishers so that they can make an explainable decision.
4. As part of the news verification process, journalists often need to validate information published on social media and they didn't feel it was supported by FANDANGO's functionality at the moment. This has been previously classified as out of the scope of the project, but it has been noted as impactful by users.

These topics will handled as top priorities and they will be assessed in the next project meeting, evaluating what's essential for the scope of FANDANGO project, so that they can be properly looked after ahead of the next pilot iteration.

It is important to highlight that the effectiveness of the project is being greatly driven by sistematic analysis and user feedback, as these results will serve as basis to define the next steps of the project and improve the quality of the implementation ahead of the second pilot iteration to be performed as D6.3. These topics will be further analyzed then, and properly documented with an updated outcome.