

# **Usability test protocol of a fake news detection platform**

<b>Author</b>	<b>Daniel Blanche-Tarragó</b>
<b>Revised</b>	<b>Andrea Rosales Climent</b>
<b>Version</b>	<b>1</b>
<b>Approved</b>	<b>11/01/2024</b>

## Table of Contents

<b>1. Introduction</b>	<b>3</b>
<b>2. Goals</b>	<b>3</b>
<b>3. Research Design</b>	<b>4</b>
3.1 Waves of Testing	4
3.2 Test Session	5
<b>4. Methods</b>	<b>5</b>
4.1 News Article	5
4.2 Prototype	6
4.3 Questionnaires	7
4.3.1 Questionnaire for Moderator(s)	7
4.3.2 Questionnaire for Participants	7
4.3.3 Sociodemographic Questionnaire	8
4.4 Translations	8
4.5 Informed Consent Form	9
4.6 Guidelines	9
4.7 User Experience (UX) Improvements	9
<b>5. Recruitment</b>	<b>10</b>
<b>6. Participants</b>	<b>10</b>
<b>7. Next Steps</b>	<b>12</b>
<b>References</b>	<b>13</b>

## 1. Introduction

Recent trends in the increasing digitalization of various aspects of social life present a staggering myriad of opportunities and possibilities, but they also pose challenges that need addressing. One such challenge is the ubiquitous and accelerated production, circulation, and consumption of disinformation, primarily taking the form of so-called fake news in recent years.

Under the framework of the **European Interest Group (EIG) CONCERT-Japan's 2020 call "ICT for Resilient, Safe and Secure Society,"** the DISSIMILAR project ([Detection of fake news on Social Media pLAtfoRms](#)) aims to develop innovative technology that helps digital users autonomously detect altered multimedia content spread through social media and instant messaging applications. The project brings together a multidisciplinary research consortium, including the Universitat Oberta de Catalunya (UOC; Spain), Warsaw University of Technology (WUT; Poland), and Tohoku and Okayama University (TU, OU; Japan). This effort is designed to remain active between 2021 and 2024.

One of the research stages included in this project is a Usability Test, which aims to assess the technology under development with potential users. This document outlines the protocol created to conduct this usability test within the three national and cultural contexts involved in the consortium.

## 2. Goals

The objective of this study was to assess the usability of the platform technology being developed. The platform will allow internet content producers to embed a watermark in their multimedia files (image, audio or video) and internet users to check whether multimedia files used on the internet have been modified since their original published version.

Leveraging the continuous work-in-progress nature of this technology throughout the project, the goal was to iteratively test the preliminary versions of the final-user platform. The rationale was that the graphic interface and functionalities could significantly benefit from the experiences of potential users, and their feedback could serve as valuable input for suggested improvements. Consequently, the focus of this usability testing was to test the design, navigability, and clarity of the platform.

More specifically, this study examined the watermark technology embedded in multimedia files analyzed by the platform. During these tests, users were tasked with determining whether the analyzed multimedia content had been altered or not, based on the information provided by the platform. The employed watermark technology facilitated addressing this question. However, it's important to note that the study did not intend to evaluate whether multimedia content creators successfully embedded a watermark in their creations.

### 3. Research Design

To achieve these goals, the team adopted a research-based user experience (UX) approach. This field focuses on improving systems through an iterative process of testing and redesign based on the experiences of external testers. Using knowledge from this area of expertise, testing was conducted on a partially developed, high-fidelity prototype (Pernice, 2016) that simulated the functionality and appearance of the technology under development. The simulative nature of the prototype was intentional on the part of the research team, but unknown to the testers, referred to in the field as the Wizard of Oz method (Ramaswamy, 2022). Testing the prototype rather than a fully developed system was considered preferable because the iterative refinements to be implemented could be done with relative ease and speed.

#### 3.1 Waves of Testing

Given the international scope of the project, the usability tests were conducted in the three countries of the consortium to take into account the cultural specificities of the potential users of the technology. Instead of conducting the individual tests with multiple participants all at once, we opted to conduct them in three iterative waves, as more insight can be gained from iterative tests with fewer participants than from a single, larger test with all participants (Nielsen, 2000). Each iteration involved collecting and analyzing feedback from participants in all three countries, making decisions about modifications to the prototype, implementing them, and debriefing the results with the consortium partners. This enabled consistent communication of the study's status among all project members. Each wave took place every three to four weeks.

The target number of participants for each wave was three in each country, totaling nine participants per country. Tests were conducted between March and May 2023 (see Table 1).

**Table 1. Time Distribution of Waves of Testing**

Wave 1	Wave 2	Wave 3
March 27, 2023 – March 31, 2023	April 17, 2023 – March 21, 2023	May 15, 2023 – May 19, 2023

In terms of the multimedia content used, the initial two waves included images, along with an audio and a video file in a Wizard of Oz approach (see Research Design). In the third wave, we tested the output of the analysis algorithm using images, since it was the only format that the algorithm could analyze at that time. Based on the technology's detection capabilities developed at the time of testing, some of the images were intentionally modified (i.e., either altered or "photoshopped," cropped, compressed, or having applied an image filter to them). The remaining images were intentionally left unmodified. These images were incorporated into both the generated news article and prototype to simulate the technology's ability to detect modified and unmodified images.

### 3.2 Test Session

The usability test involved assigning a task to a participant and evaluating how they solved it using the prototype, observing the participant's process, and assessing potential difficulties. Test sessions were conducted with different participants online and individually, moderated or facilitated by project members, and occasionally assisted by additional observers. The task simulated a situation in which participants utilized the prototype to verify whether or not the multimedia content (i.e., images, audio, and video files) of a news article they read on the internet had been altered. Participants worked autonomously (i.e., without the moderator's assistance), all while thinking aloud as the participant performed the task (Nielsen, 2012). During the test session, the moderator evaluated the participant's demonstrated and stated level of difficulty. Subsequently, the participant was asked to provide feedback on their experience with the prototype by completing an online form.

Regarding the prototype testing, the participant was instructed to complete three successive steps of testing. The primary purpose of conducting the test in three distinct steps was to familiarize the participants with the prototype's structure, functions, information, and layout. As with any new technology, users undergo a learning process that enhances their experience of using it to be more effective and efficient with continued use.

## 4. Methods

### 4.1 News Article

We created a fictitious news article incorporating multimedia content for analysis by the prototype. To simplify the testing process, the news article drew from a [genuine news story](#) (Bowman, 2023) in English, covering a relatively recent event with global coverage at that time. All of its multimedia content and an excerpt of the text were copied and pasted onto a [Blogger](#) website to imitate a real news article. Three different versions of the same news article were created, each containing different multimedia elements matching the testing steps of the prototype:

1. The first version included one single multimedia file.
2. The second version contained three multimedia files.
3. The third and final version contained six multimedia files.

The original English article was translated into the local languages by consortium partners. The news articles used during prototype testing for each step in the featured languages are listed in (see Table 2):

**Table 2. News Article Used in the Prototype Testing**

Step 1	Step 2	Step 3
<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a> <a href="#">PL</a>	<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a> <a href="#">PL</a>	<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a> <a href="#">PL</a>

*Note.* CA: Catalan; EN: English; ES: Spanish; JP: Japanese; PL: Polish. The original language version is marked in **bold**.

## 4.2 Prototype

The prototype of the platform was generated using [Wondershare Mockitt](#), a proprietary software designed for crafting software application prototypes. Since it was not a functional prototype, we simulated the user's interaction with the system through three different versions of the prototype, corresponding to the three versions of the news article to be tested. In each version, the prototype simulated the analysis of the multimedia content contained in each version of the news article. Based on the state of the technology under development at the time of user testing, the prototype comprised five successive screens.

1. The **Language** screen was the prototype's initial home screen. It presented a menu of languages in alphabetical order, including Catalan (Català), Spanish (Castellano), English, Japanese (日本語), and Polish (Polski), corresponding to those spoken by the project consortium.
2. After selecting a language, the user is directed to a second screen, **URL**, prompting them to submit the URL or link of the news article containing the multimedia files for analysis.
3. Once the user instructed the system to run the analysis, the system launched a third screen, **Analyzing**. This screen notified the user that the analysis was underway and may require some time (seconds or minutes) to complete.
4. Upon completion of the analysis, the **Results** screen appeared, displaying the results. The initial section presented a summary of the results, providing an overview of the multimedia files found in the news article, and the number of modified and unmodified files detected by the platform. The subsequent section furnished more comprehensive results concerning each multimedia file analyzed.
5. As the user navigated this screen, a functionality was presented offering enhanced outcomes or supplementary information. This ultimate fifth display, **More Information**, was created to supply further elucidation on the obtained results or explicate the technology's workings to the user.

The prototypes used in each step and wave of the prototype test, available in the featured languages, are as follows (see Table 3):

**Table 3. Prototypes Used in the Prototype Testing**

Wave	Step 1	Step 2	Step 3
Wave 1	<a href="#">CA EN ES JP PL</a>	<a href="#">CA EN ES JP PL</a>	<a href="#">CA EN ES JP PL</a>
Wave 2	<a href="#">CA EN ES JP PL</a>	<a href="#">CA EN ES JP PL</a>	<a href="#">CA EN ES JP PL</a>
Wave 3	<a href="#">CA-EN-ES-JP-PL</a>	<a href="#">CA-EN-ES-JP-PL</a>	<a href="#">CA-EN-ES-JP-PL</a>

*Note.* CA: Catalan; EN: English; ES: Spanish; JP: Japanese; PL: Polish. Wave 3 used a single multilingual prototype.

After collecting feedback from participants and project partners across all countries during each wave, we analyzed the necessary modifications to improve the prototype, and implemented them into a revised version. The updated prototype was subsequently tested in the following wave of testing.

The prototype was designed to be tested via computer or tablet, rather than smartphones. The session was held using the web conferencing platform of each partner's choice, provided they met certain criteria: (1) allowing participants to share their screen with the moderator(s), (2) providing a chat box for sharing links to the news article and the prototype, and (3) enabling the session to be recorded. Participants were only required to activate their microphone, while sharing their webcam was optional.

### 4.3 Questionnaires

We developed three questionnaires to collect data from the test: one for the moderator, one for participant feedback, and one for sociodemographic information.

#### 4.3.1 Questionnaire for Moderator(s)

The moderator(s) evaluated the difficulty of the participant's interactions with the prototype based on verbalizations and observations of level of difficulty. The assessment was conducted through completion of a [Google Form](#) survey of 10 questions addressing every step of the prototype test (e.g., accessing the prototype's website, copying and pasting a link into the prototype, navigating to the following task, understanding the results of the analyses). This assessment could be performed while the participant was carrying out the task or afterwards by reviewing the recording of the session. The questionnaires, which were only available in English, are listed in Table 4:

**Table 4. Questionnaire Addressed to the Moderator(s)\***

Wave	Step 1	Step 2	Step 3
Wave 1	<a href="#">EN</a>	<a href="#">EN</a>	<a href="#">EN</a>
Wave 2	<a href="#">EN</a>	<a href="#">EN</a>	<a href="#">EN</a>
Wave 3	<a href="#">EN</a>	<a href="#">EN</a>	<a href="#">EN</a>

*Note.* \*Documentation available to project members only. EN: English.

#### 4.3.2 Questionnaire for Participants

After completing the three steps of the prototype testing, participants were required to complete a [Google Form](#) questionnaire comprising around 20 questions. The questions pertained to their experience with the prototype, the clarity of information displayed, their views on the technology's usability in real-world scenarios, and any suggestions for improvement. The questionnaire, available in the featured languages, was presented to participants at the end of the session and can be found in Table 5:

**Table 5. Questionnaire Addressed to the Participant\***

Wave	Languages
Wave 1	<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a> <a href="#">PL</a>
Wave 2	<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a> <a href="#">PL</a>
Wave 3	<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a> <a href="#">PL</a>

*Note.* \*Documentation available to project members only. CA: Catalan; EN: English; ES: Spanish; JP: Japanese; PL: Polish.

### 4.3.3 Sociodemographic Questionnaire

In addition, sociodemographic questionnaires were distributed electronically to each participant either prior to or during the test session to collect basic profile information such as age, gender, place of birth, and educational level. The questionnaires also covered other topics such as social media usage, self-perceived digital proficiency, and subjective importance of fake news. The end-of-session questionnaire, available in the featured languages, is outlined in Table 6:

**Table 6. Sociodemographic Questionnaire Addressed to the Participant\***

Languages
<a href="#">CA</a> <a href="#">EN</a> <a href="#">ES</a> <a href="#">JP</a>

*Note.* \*Documentation available to project members only. CA: Catalan; EN: English; ES: Spanish; JP: Japanese. The Polish version of the sociodemographic questionnaire was distributed as a Microsoft Word file.

### 4.4 Translations

A collaborative [Google Sheet](#) was created to allow all partners to submit their translations of the news article, prototype, and participant questionnaire. English served as the common language, from which translations were derived for each partner's local language(s): Japanese in Japan, Polish in Poland, and Catalan and Spanish in Spain (as Catalonia, where the UOC is based, is a bilingual region). New translations had to be introduced in successive waves of testing as the instruments were modified based on feedback from participants and deliberations by the research team. The translation file is listed in Table 7:

**Table 7. Translation of Research Instruments\***

Languages
<a href="#">CA-EN-ES-JP-PL</a>

*Note.* \*Documentation available to project members only. CA: Catalan; EN: English; ES: Spanish; JP: Japanese; PL: Polish.



#### 4.5 Informed Consent Form

Finally, an informed consent form was electronically distributed in advance to each participant, who had to sign and return it to the local team in order to participate in the prototype test. The UOC provided an English model form to all partners for translation and adaptation to local conditions, which included predefined information to guide the partners. The model informed consent form, available only in English, can be found in Table 8:

**Table 8. Model of Informed Consent Form\***

Languages
<a href="#">EN</a>

*Note.* \*Documentation available to project members only. EN: English.

#### 4.6 Guidelines

We provided a guide with instructions to all partners on how to conduct the test and links to the instruments (documentation) for each location. A separate guide file was created for each wave of testing and was shared with all partners in advance (see Table 9):

**Table 9. Guidelines for the Prototype Testing\***

Wave	Languages
Wave 1	<a href="#">EN</a>
Wave 2	<a href="#">EN</a>
Wave 3	<a href="#">EN</a>

*Note.* \*Documentation available to project members only. EN: English.

#### 4.7 User Experience (UX) Improvements

After completing the prototype test in each wave, the data collected from the questionnaires addressed to moderator(s) and participants, along with feedback provided by the participants themselves during the test, were added to a [Google Sheet](#). We categorized the data into aspects of the prototype that required improvement. For every requirement, a proposal was devised, and the proposal's scope was identified either as applicable to the prototype or to the technology under development. The decision to whether implement the proposed solution or not was made based on a discussion with all the project partners. Applicable solutions were incorporated into the prototype for testing in the next wave of testing. Any suggestions falling under the technology's domain were noted as potential improvements for future consideration. In this screening process, there were also recommendations beyond the range of feasible enhancements. Those data, their categorization, and the decisions regarding implementation are presented in the file below (see Table 10):

**Table 10. Prototype Usability Improvements per Wave\***

Languages
<a href="#">EN</a>

*Note.* \*Documentation available to project members only. EN: English.

## 5. Recruitment

Each partner was responsible for the recruitment of participants in their respective country. The group of participants at each wave was expected to be diverse in terms of age, gender, and digital literacy to enable testing the prototype with a variety of potential users. Although we acknowledge that the criteria for selecting participants do not encompass the full range of potential users, we deemed them adequate for the purposes of facilitating the recruitment process. Nonetheless, it was not possible to achieve diversity in all cases. Partners were free to invite the same participants from a previous study within this project, or to invite new participants. In the case of the former, sociodemographic data collected earlier facilitated a focused invitation plan derived from their self-reported information (i.e., age, gender, self-perceived digital literacy) to categorize users for each wave. Nonetheless, to enable their participation in the second user study, new informed consent forms had to be distributed and signed. In the case of the latter, new participants were required to complete and return both the sociodemographic questionnaire and consent forms to the local moderator in order to take part in the testing.

The UOC team implemented the first strategy: all nine participants from the previous study were contacted directly by email and invited to participate in the study. Only one participant declined the invitation and a replacement with a similar sociodemographic profile had to be sought. Informal requests were extended to colleagues in the university community to reach out to their respective networks for the purpose. The WUT team recruited participants by distributing invitations throughout their social networks and extending invitations to their social circles. The TU-OU team distributed invitations through the social networks of the assisting moderators or facilitators, with an emphasis on trying to attract participants of different ages.

## 6. Participants

Twenty-five participants were involved in the study, of which sociodemographic data was collected from 23 participants. Overall, the composition of the groups across waves follows a relatively consistent pattern (see Table 11; see this [link](#) to check the file with all the results<sup>1</sup>). The groups maintained an average age in the mid-30s across waves, with less variation in Wave 2 partly due to missing data in Japan. The gender distribution shows a slight preponderance of men (13) over women (10). As expected, all groups are highly educated and predominantly employed in paid positions (often balancing work and studies), with perceived income levels similar to or higher than the average in each location.

---

<sup>1</sup> Documentation available to project members only.

When it comes to social media, all of them are social media users. They all actively engage with multiple platforms, however, preferences differ by location. YouTube is the most commonly used social media platform across all regions. In Spain and Poland, WhatsApp dominates the instant messaging category, whereas Line is the preferred platform in Japan. Instagram and Facebook were cited consistently throughout the study waves, while individuals who perceived their digital skills to be high cited Reddit more often. This subjective assessment remained relatively stable across all groups and above the midpoint of the scale. The participants' perceived significance regarding the issue of fake news appears to be substantial. However, it was predictable as they voluntarily took part in the usability study with full information about the topic, goals, and procedures of the project.

**Table 11. Characterization of the Participants**

	Wave 1	Wave 2	Wave 3	Total
<b>Age<sup>1</sup></b>	37.33 (19.10)	32.33 (8.24)	35.88 (15.45)	35.52 (15.11)
<b>Gender</b>				
<b>Woman</b>	3	5	2	10
<b>Man</b>	6	1	5	13
<b>Educational level</b>				
<b>Secondary</b>	1	1	2	4
<b>Tertiary</b>	8	5	6	19
<b>Employment status<sup>2</sup></b>				
<b>Paid work</b>	3	5	8	16
<b>Unpaid work</b>	1	1	0	2
<b>Unemployed</b>	1	0	0	1
<b>Student</b>	6	2	3	11
<b>Retired</b>	2	0	0	2
<b>Household income level</b>				
<b>Below average</b>	1	0	2	3
<b>Similar to average</b>	4	4	2	10
<b>Above average</b>	1	1	4	6
<b>Prefer not to answer</b>	3	1	0	4
<b>Social media use<sup>2</sup></b>				
<b>Facebook</b>	4	6	6	16
<b>Facebook Messenger</b>	3	4	4	11
<b>Instagram</b>	6	5	6	17
<b>Line</b>	3	0	1	4
<b>LinkedIn</b>	0	1	1	2
<b>Pinterest</b>	1	2	2	5
<b>Reddit</b>	2	0	4	6
<b>Signal</b>	3	2	0	5
<b>Snapchat</b>	0	1	0	1
<b>Telegram</b>	2	3	0	5
<b>TikTok</b>	0	2	1	3
<b>Twitter</b>	5	2	3	10
<b>WhatsApp</b>	4	6	7	17

**Table 11. Characterization of the Participants**

	Wave 1	Wave 2	Wave 3	Total
<b>YouTube</b>	8	6	8	22
<b>Self-perceived digital skills (0 – 5)<sup>1</sup></b>	3.67 (1.12)	3.33 (1.37)	3.75 (1.04)	3.61 (1.12)
<b>Importance of fake news (0 – 10)<sup>1</sup></b>	8.56 (1.33)	9.00 (1.10)	8.63 (2.39)	8.70 (1.66)
<b>N</b>	<b>9</b>	<b>6</b>	<b>8</b>	<b>23</b>
<b>Japan</b>	3	..	1 <sup>3</sup>	4
<b>Poland</b>	3	3	3	9
<b>Spain</b>	3	3	4	10

*Note.* <sup>1</sup>The number in each cell of the row represents the mean, while the number in parentheses represents the standard deviation. <sup>2</sup>The numbers in each column do not add up to total N because this was a multiple-choice question. <sup>3</sup>Three participants participated in the prototype test of this wave, but sociodemographic data were collected from only one of them. .. : No data.

Regarding the challenges faced during the tests, TU-OU in Japan did not conduct Wave 2. They resumed activities for Wave 3, however, only one participant's sociodemographic data was collected, despite three participants being involved in the prototype test. In Wave 3, the UOC in Spain conducted the test with four participants instead of the originally planned three. Due to technical difficulties, one test was not administered properly. However, the participant was permitted to complete the tasks for feedback purposes. As a result of this uncompleted test, an additional participant was recruited and included in the testing wave.

## 7. Next Steps

The assessment of the participants' performance by the moderator, the challenges and opinions reported by the participants, and the precise details of suggested improvements will be published in peer-reviewed academic outlets.

Following steps involve designing and executing a final usability test of the platform. The final version of the platform integrates the different technologies under development in the project—i.e., watermarking, media forensics, and deep learning. The main aim of this final usability study is to test the running application at the preliminary stage for testing. We anticipate that pilot testing will be conducted similarly to the prototype testing, relying on successive waves of testing with either the same or new participants.

For this final study, all suggested improvements must be applied to the technology rather than the prototype. For this, sufficient time must be allocated between waves for implementing changes. Dates for conducting these tests are yet to be discussed among the partners in the consortium, but should have a reasonable gap from the project's formal end date.

## References

- Bowman, E. (2023, January 8). *Security forces regain control after Bolsonaro supporters storm Brazil's Congress*. NPR.  
<https://www.npr.org/2023/01/08/1147757260/bolsonaro-supporters-storm-brazil-congress-lula>
- CTV News. (2023). *Supporters of ex-president Jair Bolsonaro storm Brazil's Congress*. CTV News. <https://youtu.be/Howo8ccsmwo>
- Juanjodi. (2023, January 9). *Security forces in Brazil regain control after Bolsonaro supporters storm Congress* [audio]. Podcast NPR World Story of the Day.  
[https://www.ivoox.com/security-forces-in-brazil-regain-control-after-bolsonaro-audios-mp3\\_rf\\_101159107\\_1.html](https://www.ivoox.com/security-forces-in-brazil-regain-control-after-bolsonaro-audios-mp3_rf_101159107_1.html)
- Nielsen, J. (2012). *Thinking aloud: The #1 usability tool*. Nielsen Norman Group.  
<https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>
- Nielsen, J. (2000). *Why you only need to test with 5 users*. Nielsen Norman Group.  
<https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>
- Pernice, K. (2016). *UX prototypes: Low fidelity vs. High fidelity*. Nielsen Norman Group.  
<https://www.nngroup.com/articles/ux-prototype-hi-lo-fidelity/>
- Ramaswamy, S. (2022). *The Wizard of Oz method in UX*. Nielsen Norman Group.  
<https://www.nngroup.com/articles/wizard-of-oz/>