



Baltic Engagement
Centre for Combating
Information Disorders

PROTOCOL FOR TESTING AI AND VERIFICATION TOOLS

NOVEMBER 2025

MAIA KLAASSEN | COORDINATOR

... IN CO-CREATION WITH THE WHOLE HUB AND ITS STAKEHOLDERS.

D3.2 for the EU-funded project with the ID 101226242, coordinated by the University of Tartu (UTARTU).



**Co-funded by
the European Union**

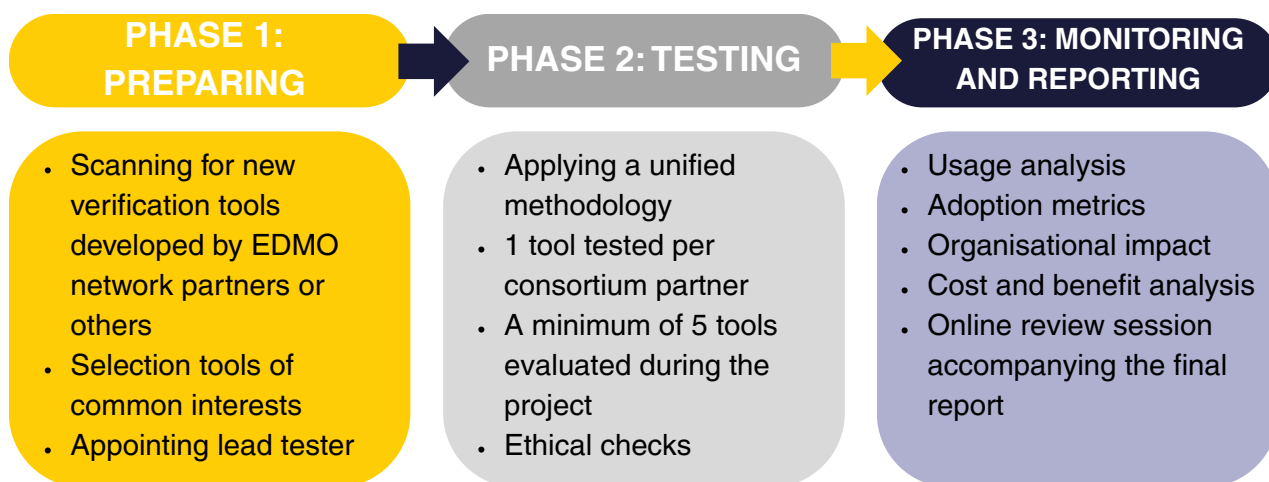
INTRODUCTION

Disinformation is not merely a technical issue, nor is it simply a communication problem; it is a societal challenge that undermines trust, polarises communities, and threatens democratic processes. As one of the 15 **EDMO** hubs working together across the EU (and in Norway, Moldova, and Ukraine), BECID2 aims to build on the achievements of its predecessor, strengthening societal resilience against disinformation through fact-checking, media literacy, and policy monitoring. As a regional hub spanning three countries and featuring three different state languages (Estonian, Latvian, and Lithuanian), with Russian and English serving as additional working languages for this project, we address the growing complexity of the information ecosystem. This complexity is compounded by AI-generated content and sophisticated influence operations, which challenge traditional verification methods.

With the increasing amount of AI-generated content, it is unreasonable to expect debunking to be done solely by human hands. On the other hand, as research from BECID1 has found, there is an overload of new verification or debunking tools launched or in development, with very few standardised tools actually changing the work processes of journalists, fact-checkers, researchers, or analysts (other than Large Language Models) so far.

THE ABUNDANCE OF THE TOOLS THAT PROMISE TO EASE THE WORKLOAD IS SO GREAT THAT THE PROCESS OF FINDING, TESTING AND INCORPORATING THESE TOOLS HAS BECOME TOO TIME-CONSUMING.

Therefore, we developed a collaborative task (T3.3) for the testing of new tools. BECID2's approach is holistic, combining investigative rigour with technological innovation, ensuring that human judgment remains central while leveraging some level of automation for scale and efficiency. Rather than a static approach, we aim for this protocol to evolve through iterative testing, feedback loops, and collaborative refinement across the EDMO network.



The first preparatory phase of the protocol focuses on continuous environmental scanning and strategic selection of tools, ensuring that testing efforts are aligned with real-world needs and consortium priorities.

PHASE 1: PREPARING

Discovery	Environmental scanning by the partners to monitor developments in AI and OSINT verification tools, including those emerging from EDMO hubs and the private sector.
Consensus building	The consortium identifies tools of common interest through first , initiating a shared table for listing down tools they have seen, read about, heard being used in the EDMO network in other professional networks prior to the initial meeting. Second , a meeting with a structured discussion will be held, where the list of tools will be assessed and interest expressed by fact-checkers and researchers. A final list of what we shall test will be confirmed, and testers assigned to each tool with a deadline and progress monitoring in our Notion dashboard. Third, adaptability for the results of the feedback loop of our consortium governance will be maintained. Updates given during Steering Committee meetings on an ad hoc basis will be incorporated in the tool list as needed.
Tool selection	Selection criteria: • Tools selected must address disinformation detection or verification challenges relevant to the Baltic context. • Explicit checks for multilingual capability (Estonian, Latvian, Lithuanian and Russian in addition to English). • Bias risk pre-assessment to identify potential language or demographic bias risks during tool selection. • Ethical compliance and security pre-screening. Foreseen risk lists with proposed mitigation strategies will be prepared for tools where necessary (such as Deepseek, for example, which could be implemented

	via a local model version, but should have the source code changed to the underlying model) as per our ethics principles.
Lead Tester appointment	For each selected tool, a lead tester from one of the BECID partners is designated to coordinate evaluation activities.

The second phase ensures that testing is conducted under a shared methodological framework to guarantee comparability and rigour, transforming abstract interest into actionable evidence, and therefore enabling informed decisions on adoption.

PHASE 2: TESTING

Scope	At least five tools will be tested during the 30-month period, with each partner responsible for one tool.
Ethical and Security checks	Ensure compliance with BECID's Ethics Report, GDPR, and EDMO network standards throughout testing. Conduct privacy risk assessment and, where relevant, penetration testing for cloud-based tools.
Unified methodology	Standardised Steps: • Establish a benchmark: baseline definition of current workflows and performance metrics (time, accuracy) before introducing new tools. • Define the tool's intended function and context of use. • Conduct initial usability assessment (interface, onboarding, scalability with high-volume tasks). • Bias & fairness testing across multiple languages and content types before uploading any data. • Assess integration feasibility and compatibility with existing (organisational and pan-consortium) CMS, fact-checking platforms, and workflows. • Apply the tool to a real verification task or dataset. • Document performance metrics (accuracy, speed, error rates, precision, recall, F1-score, false positive/negative rates, etc). • Capture qualitative feedback from testers.

	• Structured UX evaluation: where relevant, apply a standardised usability scoring system (e.g., SUS – System Usability Scale) alongside qualitative feedback. • Fill out Tool Evaluation Form prepared by the coordinator. Turning abstract interest into actionable evidence: final recommendation on adoption presented in a online session with invites sent to the EDMO network, announced publicly and open to anyone interested.
--	---

The third phase evaluates the testing, going beyond technical performance to address organisational and strategic dimensions, ensuring accountability and providing a robust evidence base for strategic decisions on tool integration.

PHASE 3: MONITORING & REPORTING

Usage overview	• How was the tool selected? • What was the usage duration (a day, a week, 6 months)? • Was the tool used for a one-time task or integrated into workflows? Describe the task.
Adoption metrics	• How many people tested the tool? • After the testing period, how many indicated willingness to continue using it?
Diffusion beyond BECID and motivation	• Who tested the tools in the organisation? Did interest extend to individuals or teams outside the consortium, signalling broader organisational or community relevance? • What was the motivation for the adoption of this tool? Was adoption driven by project obligations (mandated by the grant agreement) or genuine interest?
Cost & ROI	• What is the price of the tool? • How does its performance justify the investment? ROI analysis based on qualitative feedback from testers.

	• How do the post-testing metrics compare against the baseline established in Phase 1? • What is the final recommendation on adoption?
Reporting outputs	• Circulating the completed tool evaluation form along with the invitation to an accompanying online review session in the consortium, to other EDMO partners and within our own networks. • Based on these evaluation forms, three monitoring reports presented to the EU (MS20) by M32, synthesising findings and recommendations. For the monitoring reports, a post-testing audit will be scheduled to evaluate sustained impact and continued use.