**DisinfoRadar Research Onboarding**

|  |
| --- |
| **Research/technical outputs:**  #1 Threat Registry (includes strategy, content, timeline, recommended team distribution)  #2 Deep Dive Report  #3 BTW Meta-Report  #4 Rapid Response Briefs  #5 Bi-Annual DisinfoRadar reports  *The document is based on handover documents from former colleagues, Research Coordinator and Programme Officer.* |

**#1 DisinfoRadar Threat Registry**

What will be crawling (scope, geo)?

This task will benefit from a brainstorming session based on the following framework:

Geographically there is no scope, though we will be constrained by language. FO/AA does have special interest in Russia and China though so we can skew our collection to ensure we capture these. More generally special attention can be given to countries where the tools are being developed, i.e. usually technological advanced countries, and countries where new tactics are being used, i.e. usually democracies backsliding or still unconsolidated. On language, there are translation algorithms.

This segues to topically scope, which has two aspects, the tools and the tactics. New tools can be gathered from scientific publications (e.g. IEEE and its numerous publications, including some rather accessible). Note that Chinese abstracts tend to be in English, but tactics could be harder.

Tactics is harder and may require gathering from news sources. Search will have to be broad though, as these local non-experts may not know what they are looking at (and that is our value add: simplifying, we search “fake news” and give context and depth).

For both tactics and tools, we should also monitor groups devoted to this topic (e.g. Reddit, and perhaps Twitter) and expert blogs, as these allow more imagination and speculation. Another option is to monitor known disinformation actors (e.g. RT) and watch how they distribute their content (e.g. the new trend of using influencers, not just bots).

What actually needs to be built technically speaking (by the web developer v. data science expert)?

The following are the main technical components to be built:

1. A web scraper, web crawler, or combination of the two:
   1. In most cases a crawler will be more useful here, to crawl through known websites looking for examples of tools or tactics. Scraping generally has to be more targeted, though may be useful for groups, blogs, etc.
2. A pre-processing script to turn the collected results into data we can analyse
   1. Recommend also developing means of putting collected results into an SQL or No-SQL database. This will facilitate web deployment and also is a good backup in case the algorithms for sorting and filtering do not perform well (more on this below).
3. Text summarization algorithm and script (suggestion from former RC: this could be dropped)
   1. Depending on the type of content the crawler returns, this may become a lower priority item, possibly even a task we can drop (though as a super-goal for, e.g. version 2.0, this would still be ideal).
   2. Lower hanging fruit here would be rather than full summarization we extract keywords (nouns, dates, etc.). This can also serve as a simple, backup means of performing task four.
4. Text clustering and/or topic modelling algorithm and script, or other means of anomaly detection
   1. See also point 3.b. above. Please note though, 1) clustering algorithms are not that difficult and 2) the backup plan is truly a back-up, as in sub-optimal.
5. Decision point related to a data engineering question: how do we want to get the findings from data onto the website?
   1. Minimum viable product: a manual process of writing up content and inserting into a web framework capable of having continuously updated content.
   2. Goal: Outliers / anomalies would, with summaries of texts as required, automatically populate website
   3. Super goal: algorithm can combine tools and tactics and generate hypotheses on application (e.g. think of how one feeds texts to GPT3 and gets believable texts back).
6. Web platform to display findings
   1. This needs to be dynamic, more like a blog than the DRI toolkit, as we cannot know in advanced how many texts, or how long they will be, end up on the website.
   2. Aside from this dynamic nature, it should be elegant, simple, and easy to navigate and to view combinations or apply filtering

In this system, how will we know that we have found something worth reporting on?

See section 2.2, though for the registry less importance given to 'urgency'.

How will this information be summarized/consolidated to actually put on the platform (assuming there will be some filtering process with writing)? Constant monitoring? Some layer of analysis?

See above on what we will crawl, in combination with following points on potential issues.

Researchers will hopefully only have to look at the list and clustered/summarized data to identify something strange and unusual. If they have to review all of the content manually, that would take much longer.

How will the platform content be organized?

See mock-up slides [here](https://democracyreporting.sharepoint.com/:b:/r/sites/SocialMedia/Shared%20Documents/3_Fundraising/AA%20Disinformation%20Radar/Charts%20and%20images/Registry_mockup.pdf?csf=1&web=1&e=vmPWUo). Early consultation with web-developer important here to know what is feasible within time and budget constraints.

Potential issues future people should consider and thoughts on Plan B? What if tool does not pick up something interesting? Could the work be done qualitatively?

If the clustering is failing, we will notice if it’s garbage/fails right after starting.

Should the algorithms not work effectively I recommend basic information aggregating and filtering via Python and/or SQL queries (keyword searches) to build the tools and tactics sections. To be colloquial, there is a date-science adage “Neural network for the resume, SQL query for production”. Do not be shy to use the later, and if your algorithms feed into the SQL table, then you have actually done very well.

The ‘story’ section was designed to be more qualitative, even if the aforementioned super-goal would be automation. Note also, the inclusion of ‘story’ element was in part done to give some flexibility on outputs. The author just needs to be creative!

If we end up having to go with a keyword approach, this will be a problem because we will only find what we are looking for. We will need to then come up with a list of common but important words that could flag problems. It will require some brainstorming. Example searching for “fake news” or “deep fake”. We could also try clustering within articles/content mentioning “fake news” to identify new narratives spreading globally.

Recommendations to inform implementation

Note that a web developer and data science skillset is different, as is a data engineer (who often is needed to for the interface between data science and web development)

Timeline and sequencing

Time estimates included (under duress) and should be considered very rough – anticipating complications, skill level of those working on task, other tasks breaking flow, makes prediction unreliable.

Time in:

1. Define collection criteria
   1. Time to complete (TTC): 2 weeks, preparatory research, concept meeting, code tests (ideally – feasibility), pitch meeting (how code went and recommendations), decision
2. Design Consultations
   1. On web site design to inform our collection and analysis
      1. TTC: 2-3 days of iterative meetings, drafting, and discussion. Can be longer if some initial design and integration decisions are made here.
   2. On overall project design:
      1. TTC: 2 weeks iterative meetings, research, recommend flow charts, full code work plan, extra attention to integration of all components.
3. Build web crawlers and scrapers
   1. TTC: 1-2 weeks
4. Examine data returned, revises crawler and scrapers
   1. TTC: 1-2 week, in conjunction also with task 5
5. Build a pre-processing script to turn the collected results into data we can analyse and an SQL database
   1. TTC: 2 weeks, first week contemporaneous to task 4.
6. Devise storage solution for data that integrates well into website (can include thus decisions on web hosting)
   1. TTC: A few days iterative meeting and exploration; see task 10, the initial four points can be covered here in a cursory or in-depth manner. The more in depth, the less time needed for task 10 (or the more certain more overrun time is not needed.
7. Begin website build
   1. TTC: one month, 1.5 months better (for alpha version)
8. Write code for text summarization
   1. TTC: 1-2 weeks
9. Write code for text clustering and/or topic modelling for outlier and anomaly detection
   1. TTC: 1-2 weeks
10. Based on quality of clustering and summarization determine data engineering question: how do we want to get the findings from data onto the website? (see section 1.1)
    1. TTC: 1 day conceptualization, meeting, deciding, 1 day option exploration, 1 week to build, can be reduced with more time spent on Task 6.
11. Internal testing of website:
    1. TTC: 2-days
12. Begin content creation, deploy to alpha version of website
    1. TTC:
       1. 2 weeks content prep before deployment then ongoing
       2. Alpha deployed 1 month [or 1.5 better] and 1 week after web design commenced (task 7)
13. Deploy beta version of website (beta = feature complete)
    1. TTC: 1 months after deployment of alpha
14. Launch version 1.0:
    1. TTC : 1 month after deployment of beta at DisinfoCon

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **YEAR** | **2021** | | **2022** | | | | | | | | | | | |
| **MONTH** | **11** | **12** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| Planning and ramp-up | **1. Define collection criteria** |  |  |  |  | **⟐** |  |  |  |  |  |  |  |  |  |
| **2.Consultations** |  |  |  |  |  | **⟐** |  |  |  |  |  |  |  |  |
| Build framework components | ***3. Build collection crawlers*** |  |  |  |  |  | **⟐** | **+** |  |  |  |  |  |  |  |
| ***4. Explore data, revise code*** |  |  |  |  |  |  | **⟐** |  |  |  |  |  |  |  |
| ***5. Pre-process script + SQL*** |  |  |  |  |  |  | **⟐** |  |  | **+** |  |  |  |  |
| **6. Storage solution** |  |  |  |  |  |  | **⟐** |  | **+** | **+** |  |  |  |  |
| **7.Web build** |  |  |  |  |  | **-** | **-** | **⟐** | **I** | **+** | **+** | **+** |  |  |
| Build analytical components + Integrate | ***8. Text summarization code*** |  |  |  |  |  |  | **-** | **⟐** | **I** |  |  |  |  |  |
| ***9. Text clustering code*** |  |  |  |  |  |  | **-** | **⟐** | I |  |  |  |  |  |
| **10. Data engineering** |  |  |  |  |  |  |  |  | **⟐ I** | **+** | **+** | **+** |  |  |
| Deployment, testing and further development | **11. Pre-Alpha testing** |  |  |  |  |  |  |  |  | **⟐** |  |  |  |  |  |
| **12. Fill content and Deploy Alpha** |  |  |  |  |  |  |  |  | **-** | **⟐** |  |  |  |  |
| **13. Develop - Deploy Beta** |  |  |  |  |  |  |  |  |  |  | **⟐** | **+** |  |  |
| **14. Launch 1.0 web** |  |  |  |  |  |  |  |  |  |  |  | **⟐** |  |  |

**I** = Integration (recommend full attention, hard and heavy task possible)

**+** = task continues in other form

***Italics*** = data science code task (likely) requiring ongoing revision and updates to improve quality

Deadline

October 2022.

Recommended staffing modality

* Research coordinator: Workstream package lead, drafting conceptual documents and outlining technical/content structure, coordination with experts to execute platform build (delegate oversight of specific components of build to Programme Officer (Lena) /Research Associate) TBD), drafting text for platform, review content drafted by PO/RA for website
* Programme officer (Sorina): Hiring external experts, checking in with RC that deliverables are on track to ensure implementation of other project activities are on track, coordinate specific platform build deliverables as assigned by RC
* Data science expert (TBH): Building web collection tool, advise on data engineering questions (storage, transfer, etc) and revise collection scripts in accordance with the answer
* Web developer expert (TBH): Building website, consult on design decisions
* Research Associate: Alongside RC draft content for website (and lead on static content, e.g. “about” sections), under supervision of RC make design and visualization proposals (with web developer’s involvement); in these capacities RA acts as understudy to RC (for cover and oversight of registry upon completion)

**#2 Deep Dive Report**

Ideas for report topics?

The topic of these reports will be dictated by the interests of DRI management as driven ideally by the findings of the threat registry.

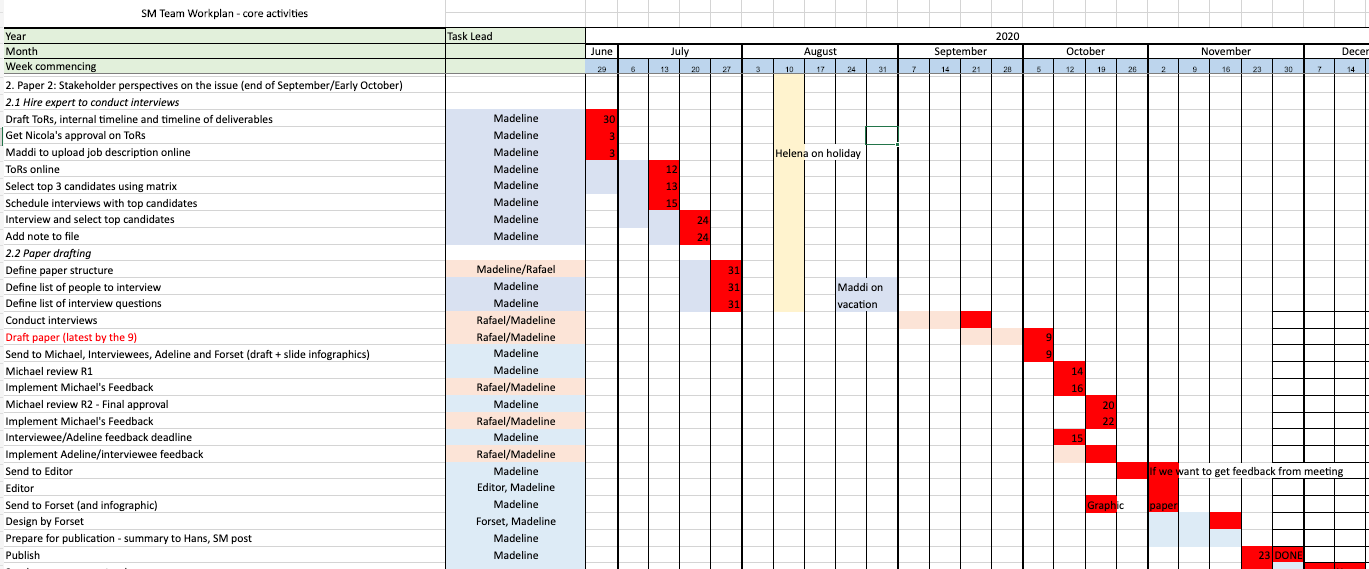
Initial example ideas pitched (titles are just examples and unlikely to be approved) included:

* GANations - Outlining the experience of states as targets of disinformation or examining competing disinformation tactics, tools, policies of different states (with a view here to balance and non-attribution, i.e. non-political)
* Deep Fakes II: Deeper and Faker – a report looking at application of GAN to domains beyond fake audio-visual content, such as generating fake data, audio

Recommendations to inform implementation

See project documents from [1085 (deepfakes)](https://democracyreporting.sharepoint.com/sites/SocialMedia/Lists/SocialMedia_Overview_Documents_Libraries/DispForm.aspx?ID=4&e=LBBAmw) particularly related to Report #2, for sample timeline structure and templates to structure work. In this project, we conducted 22 semi-structured interviews from experts for civil society (think tanks, academia), social media platforms, technology creators and policymakers to assess the potential impact/preparedness of deepfake technology. For each stakeholder group, we constructed a different set of questions. We added specific questions per interlocuter based on their research/work. Each interview lasted 1 hour except 2 interviews were conducted via email (with hard to reach individuals). To identify the interlocuters, we had a list of priority top 20 interviewees then a second/third tier of 30 and 40 if individuals from the priority list did not respond. Results were structured into an Excel spreadsheet per question and a rough summary report was drafted. These two documents fed into the drafting of the final report. An external expert supported on identifying interlocuters, drafting interview questions, organizing interviews and drafting a rough summary report. For interviews, we did not take exact notes but drafted summary of each answer with a top-level summary per interview. We told interviewees that their org would be named but not their title/name specifically so they would speak more freely.

*See* [*2020 Workplan*](https://democracyreporting.sharepoint.com/:x:/r/sites/SocialMedia/_layouts/15/Doc.aspx?sourcedoc=%7BD3B16F44-7629-4A9C-9872-FAB4ED985ED5%7D&file=SM%20Team%20workplan%202020.xlsx&action=default&mobileredirect=true) *for how work was structured:*



If topics are more niche than deepfakes, you will not need to conduct interviews with 22 individuals like we did for the deepfake report. This will reduce the complexity of implementation.

Working in close contact with the GMFA on sharing the outputs is very beneficial. They can share with their networks to increase readership, invite us to events and possibly share our event invites with their contacts in the policymaking world.

When sending report to designer, already include concrete ideas for how tables should be visualized for better results. Some ugly, rough mockups could be created in PowerPoint. Although, editor would like to see raw, unformatted text so save these comments for after.

Recommendations and timeline

Overall Timeline (3.5 to 4 months – noting no external expert hired like 1085):

* + 1-2 weeks for drafting concept, getting internal approval
  + 1-2 weeks for scoping interlocuters, drafting interview questions and getting internal approval
  + 2-3 weeks for interview reach out and carrying out interviews w/ notes, in the meanwhile filling out results table and drafting interview summaries
  + I would give at least 1.5 months for this task. 3 weeks to draft the publication was absolutely too tight.
  + Remember to plan for 1 month for reviews, editing and layout/design. 1 month and 1 week if you expect 2 rounds of management review.

Recommended staffing modality

* Research Coordinator and Programme Officer (Lena) with support from Research Associate
  + Draft concept, interview questions (or PA with close review), lead on interviews, key lead on drafting final report with inputs on sections by PA
* Research associate: Drafts interlocuter list, organizes interviews, notes and interview summaries, drafts specific sections of report, supports on launch event implementation, supports comms tasks
* Programme Officer (Sorina): Project coordination, timeline management, launch event management, report editing/design liaison, MEL. Supported by Working Student.

Deadline

June 2022.

**#3 BTW21 Meta-Paper**

Numerous organizations analysed social media during the German elections to parliament in 2021. This paper synthesizes these findings to serve as an overview of the main findings of the many reports. We are interested in linkages between different areas of assessment and overarching patterns that indicate a trend for potential future threats to electoral outcomes. This research is therefore guided by the following research question: To what extent do the results of the analyses, divided by specific areas of focus, overlap in patterns that are posing a threat to electoral outcomes?

Early concept: [BTW21 early layout & research questions.docx (sharepoint.com)](https://democracyreporting.sharepoint.com/:w:/r/sites/DisinforRadar/_layouts/15/Doc.aspx?sourcedoc=%7BF8F0AD71-42F9-432B-AB83-C81B46E5AE63%7D&file=BTW21%20early%20layout%20%26%20research%20questions.docx&action=default&mobileredirect=true)

Deadline:

End of March

Staffing modality:

External expert Pollytix was hired to draft the report, under the supervision of Programme Officer (Lena- Maria B.) and Executive Director (Michael M-R)

**#4 Rapid Response Briefs**

Content, length, target audience, dissemination?

Rapid reports are intended to be circa 1-page briefings on issues of importance, in that a tool-tactic combination poses a unique and new threat to an existing vulnerability in our socio-political or technical system and the probability of this threat manifesting on a short-time line is high (short = political and policy maker time-horizon, i.e. 1 to 4 years with preference or 1 to 2). The audience is intended as the policy making community and expert community/organization working in the technology used in the threat. Reports should be produced within 1-week of threat identification (while not strictly necessary, this is a good exercise to prepare for scaling and fully capable threat registry... that and this was promised in the proposal).

What could the process from DisinfoRadar platform to deciding to draft a publication look like?

See section 2.2.

Recommendations to inform implementation

There are not too many lessons learned since this would be a new product for DRI. Although, note that the concept proposed here was really 1-2 page results, so it should not be overcomplicated. The exact target audience and dissemination of these rapid reports needs closer attention and definition.

Process for identifying topics for rapid reports entails two criteria, originality and urgency:

Originality (i.e. is the threat new?): should be assessed by whether the data is an outlier, which can be assessed using K-means clustering or topic modelling. If these algorithms are not ready text summarization algorithms can be used to produce summaries which staff can review and assess qualitatively. Random samples should be reviewed in any case. If none of these options are ready, a simple filtering script or SQL query is a last resort (e.g. new only, drop if known keywords present, importance ranking based on metrics like impact factor, etc.).

Urgency (i.e. do we need to act now?): qualitative assessment based on existence of vulnerability the new tool or tactics could target, state and complexity of tool development and deployment, and – most important – own assessment of possible combination of object (tool or tactic) with the other object (tactic or tool).

Recommended staffing modality

* Research Coordinator: develop code clustering, topic modeling and text summarization along with DS expert, review outputs for refinement of algorithms and final report topic selection; lead on research for urgency aspect, draft report
* Programme Officer (Sorina): Project coordination, timeline management, report editing/design liaison, MEL. Supported by Working Student.
* Research Associate(s): Review and label data to contribute to algorithm fine-tuning, keep a global overview of collection (i.e. read summaries) to flag up to RC, visualizations for rapid report as needed.

Note: Data science expert would not be involved in the writing of these reports. They may only be involved in writing code for clustering, topic modeling and text summarization, or for the backup plan as required (Python filtering and SQL queries).

**#5 Bi-annual DisinfoRadar reports**

Content description

First Bi-annual report is a top-level summary of new disinformation threats based on desk research and our proposed way of conceptualizing these and foreseeing them. This report should feed into development of DisinfoRadar platform.

Second bi-annual report should summarize what has been found via the DisinfoRadar platform so far, so it can be distributed with the platform launch and findings can be presented at DisinfoCon.

Mapping the disinfo scene to set the scene: If we were already starting today, what resources would we already consult? How could this report be structured?

The goal of bi-annual summary report number one was to set the framework for how we are examining disinformation; it is to be a concept piece that shifts the narrative from current threats and its simplistic presentation to one encompassing a careful consideration of tools and tactics and how these are used to create a story, what one organization has called “adversarial narratives”. Thus this report should break down threats to the information environment (dis-, mis-, etc.) into components (technical and also with a good risk assessment framework) and present our nuanced way of viewing the threat.

As a secondary aspect the report will have to discuss how we plan on “predicting the future” since we promise not only to look on the horizon (emerging tools, tactics, threats) but also over the horizon. Processes for forecasting or futurology should be reviewed and selected based on team’s expertise and preferences. For example, the previous DD coordinated preferred to use technological adoption models and path-dependency, as well as model of tactical preferences in light of organizational goals and constraints. This preference is expressed in the project design, in that by looking at tools and tactics we can predict threats (provided an actor exists that can adopt the tools and tactic and a vulnerability exists in an adversary against which the actor can target its innovation).

Summary of DisinfoRadar platform findings: Ideas on structure?

This output can be more of a traditional NGO style, project closing report if so desired.

Staffing modality

* Research Coordinator and / or Programme Officer (Lena): Workstream lead, Draft structure and concept. With support from Research Associate.
* Important that RC has the inputs from Bi-annual #1 report to conceptualize DisinfoRadar platform.
* Programme Officer (Sorina): Project coordination, timeline management, launch event management, report editing/design liaison, MEL. Supported by Working Student.

Deadline

April, end of September 2022.