| Situation | Current surveillance system is not suitable to be implemented in LMICs, especially due to lack of resource. Being aware that surveillance system to collect data regarding AMR is important | Aims | Building a prototype model which works to mine data from news available in the internet regarding antimicrobial resistance |
| --- | --- | --- | --- |

| Inputs and activities | Outputs | Change mechanism | Outcomes | Impacts |
| --- | --- | --- | --- | --- |
| **Inputs**  **What financial outlay, staffing and other resources will be required?**   * + Establish a team with expertise in web scraping, data extraction, and AMR-related issues.   + Procure necessary resources including web scraping tools, computing infrastructure, and access to online sources.   + Identify and select relevant online sources including news websites, healthcare forums, government reports, and scientific journals. | **What tangible results, products, lessons, inspections or improvements will be produced?**   * + Database containing structured information on AMR incidents including type of threat, location, affected population, antibiotics administered, etc. | **What actions will be needed to achieve the change(s)? Are you removing frictions, changing behaviour etc.?**   * Reducing the barrier and resources in accessing the available text data regarding AMR. | **Short term**  **What will be the benefits and wider outcomes, both leading and lagging?**   * + Enhanced understanding of the epidemiology and dynamics of AMR within the Indian context.   + Facilitated data-driven decision-making for policymakers, healthcare professionals, and researchers involved in combating AMR. | **What are the impacts and how do they fit with departmental and governmental priorities?**   * + Strengthened global efforts in addressing AMR by contributing comprehensive and accurate data from India to the global surveillance system.   + Informed policy interventions and public health strategies aimed at mitigating the impact of AMR on human health and healthcare systems.   + Reduced morbidity, mortality, and economic burden associated with AMR through targeted interventions and resource allocation informed by surveillance data. |
| **Activities** **What will be delivered, such as training or guidance?**   * + Develop and refine web scraping algorithms tailored for extracting AMR-related stories from the identified online sources.   + Implement quality control measures to ensure the accuracy and reliability of extracted data.   + Regularly update and adapt the web scraping algorithms to accommodate changes in online sources and reporting trends.   + Establish a storage system for collected data |  |  | **Long term**  **What will be the sustainable and lasting changes, and what metrics will be used  to measure these?**   * + Improved surveillance and monitoring of AMR-related incidents in India through timely identification and documentation of relevant stories. |  |

| **Evidence assessment** | **What is the strength of the existing evidence base for this change?** |
| --- | --- |

| **Assumptions** | **What is being assumed  as part of the plan?**   1. Availability of online sources reporting on healthcare and AMR-related issues in India. 2. Accessibility to web scraping tools and technologies. 3. Reliability of data extraction algorithms in identifying and extracting relevant information accurately. 4. Regular updates and maintenance of the global AMR surveillance spreadsheet. 5. Regular time to publish articles across geographical area 6. Every places have the same likelihood for an AMR incidence reported in the news | **Possible unintended consequences** | **Are there any other outcomes that might result from this project?**   * The query keyword maybe used differently for surveillance in different cases |
| --- | --- | --- | --- |

**Theory of Change for Web Scraping Project on AMR-Related Stories in India**

**Reseach Question:**

How can large language models (LLMs) be leveraged to improve our knowledge of antimicrobial resistance (AMR)?

**Goal:**

To establish a robust system for identifying and extracting AMR-related stories from online sources in India, extracting key information such as the type of threat, location, number of people affected, antibiotics administered, etc., and storing this information in a accessible AMR surveillance spreadsheet.

**Assumptions:**

1. Availability of online sources reporting on healthcare and AMR-related issues in India.
2. Accessibility to web scraping tools and technologies.
3. Reliability of data extraction algorithms in identifying and extracting relevant information accurately.
4. Regular updates and maintenance of the global AMR surveillance spreadsheet.

**Theory of Change:**

* ***Input:***
  + Establish a team with expertise in web scraping, data extraction, and AMR-related issues.
  + Procure necessary resources including web scraping tools, computing infrastructure, and access to online sources.
  + Identify and select relevant online sources including news websites, healthcare forums, government reports, and scientific journals.
* Activities:
  + Develop and refine web scraping algorithms tailored for extracting AMR-related stories from the identified online sources.
  + Implement quality control measures to ensure the accuracy and reliability of extracted data.
  + Regularly update and adapt the web scraping algorithms to accommodate changes in online sources and reporting trends.
  + Establish a secure storage system for collected data and integrate it with the global AMR surveillance spreadsheet.
* Outputs:
  + Automated web scraping system capable of identifying and extracting AMR-related stories from various online sources.
  + Database containing structured information on AMR incidents including type of threat, location, affected population, antibiotics administered, etc.
  + Integration of extracted data into the global AMR surveillance spreadsheet, ensuring compatibility and accessibility for stakeholders.
* Outcomes:
  + Improved surveillance and monitoring of AMR-related incidents in India through timely identification and documentation of relevant stories.
  + Enhanced understanding of the epidemiology and dynamics of AMR within the Indian context.
  + Facilitated data-driven decision-making for policymakers, healthcare professionals, and researchers involved in combating AMR.
* Impact:
  + Strengthened global efforts in addressing AMR by contributing comprehensive and accurate data from India to the global surveillance system.
  + Informed policy interventions and public health strategies aimed at mitigating the impact of AMR on human health and healthcare systems.
  + Reduced morbidity, mortality, and economic burden associated with AMR through targeted interventions and resource allocation informed by surveillance data.

**Monitoring and Evaluation:**

* Regular assessment of the web scraping system's performance in terms of accuracy, coverage, and timeliness.
* Periodic reviews of the quality and relevance of extracted data.
* Feedback mechanisms from stakeholders including policymakers, healthcare providers, and researchers to evaluate the utility and effectiveness of the surveillance system in supporting AMR-related initiatives.
* Adjustments and improvements to the system based on monitoring and evaluation findings to ensure continued effectiveness and relevance.