Plan Overview

A Data Management Plan created using DMPonline

Title: Network on Neurosymbolic Systems for Medicine

Creator: Ernesto Jimenez-Ruiz

Principal Investigator: Janna Hastings, Ernesto Jimenez Ruiz

Affiliation: City, University of London

Template: DCC Template

Project abstract:

Artificial intelligence has many applications in medicine. Recently, systems such as ChatGPT have made headlines for their groundbreaking performance on a wide range of tasks. However, these systems can make errors or act harmfully in certain contexts. The technology is powerful but in its current form of learning from data suffers from limitations. New approaches are needed which can render AI transparent and controllable while keeping the powerful potentialities. A relatively new approach, known as ‘Neurosymbolic AI’, offers methods that are able to learn from data but remain transparent and controllable via rules that can be expressed in terms of clinical knowledge. This is achieved by combining data-driven approaches with knowledge-driven approaches to develop new forms of AI that have greater potential for success in medicine.

We aim to found a wide-ranging network to establish this approach to AI for the field of medicine. Together with our core collaborators, we will plan and hold a workshop with the key researchers in neurosymbolic AI worldwide in order to define a research agenda and roadmap to establish this technology. This will have far reaching impact in leading to safer and more transparent AI for medical applications.

Funder: Academy of Medical Sciences

ID: 137481

Start date: 10-03-2024

End date: 09-03-2025

Last modified: 08-11-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Network on Neurosymbolic Systems for Medicine

Data Collection

What data will you collect or create?

We plan to gather relevant use cases and datasets for the network. Data can be multimodal, e.g., images, audio, text, video, tables. On the first iteration data is expected to be on GB range.

How will the data be collected or created?

The Network aims at gathering datasets that have been already collected. If licences permit we plan to share the gathered resources in a public GitHub repository (a distributed version control system: https://github.com/). Depending on their license, other resources may be kept private for use within the network.

Documentation and Metadata

What documentation and metadata will accompany the data?

It may be necessary to annotate the gathered resources if they lack sufficient metadata. In this case we will seek for best practices including the use of state-of-the-art ontologies and knowledge graphs for the annotation.

Ethics and Legal Compliance

How will you manage any ethical issues?

We do not foresee any ethical issues as the network will only gather already created data. If we are in doubt we will seek for ethical advise to City's SST ethics committee.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

We plan to make accessible to the network and broader community only resources that have open copyrights (e.g., CC BY 4).

Storage and Backup

How will the data be stored and backed up during the research?

Data and other resources will be added to a GitHub repository.

How will you manage access and security?

If required some resources will be made private within the GitHub repository.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

N/A

What is the long-term preservation plan for the dataset?

We plan to keep the GitHub repository active beyond the Network duration.

Data Sharing

How will you share the data?

Data and other resources will be added to a GitHub repository. In addition, for very large datasets we may consider using Zenodo (https://zenodo.org/).We will also consider the use of Haggingface (https://huggingface.co/) to share AI models.

Are any restrictions on data sharing required?

N/A

Responsibilities and Resources

Who will be responsible for data management?

The application lead and co-applicant: Janna Hastings and Ernesto Jimenez-Ruiz.

What resources will you require to deliver your plan?

N/A