Plan Overview

A Data Management Plan created using DMPonline.be

Title: Intelligent Prognosis of Rotating Machines in Industry 4.0 using Generative AI

Creator: Tauheed Mian

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Project abstract:

Rotating machines are an important component of any manufacturing system and power generation system. The safe running of these machines is to be ensured by continuous and proper monitoring. Further, the competitive market and rapid development in computational technologies have created the necessity for this system to be economical and smart. So, this makes the presence of an efficient, economical, and smart diagnosis and prognosis system mandatory. The data-driven approaches currently applied for the prognosis face the challenges of data availability, data scarcity, scalability, and many others. Generative Artificial Intelligence (GenAI) recently created a bowl to mitigate these challenges. Further, the advancement in technology related to Cyber-Physical Systems (CPS) and the Internet of Things (IoT) to have the system economical and smart enough for real-time prognosis and to make decisions accordingly. Considering these requirements of the present industrial system, this project presents the approach involving the integration of GenAI, CPS, and IoT. The outcome of this project will be utilizing the interconnected trinity of GenAI, CPS, and the IoT to ensure the longevity and efficiency of rotating machines, a more reliable industrial landscape, and mitigating the big hurdle of data scarcity.

ID: 213080

Start date: 27-11-2024

End date: 31-10-2025

Last modified: 26-02-2025

Intelligent Prognosis of Rotating Machines in Industry 4.0 using Generative AI

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Experimental data	Data of healthy and faulty conditions	E and N	D	N	.csv; .xls	<100GB	
Experimental results	interpreted data	N	D	N	.xls(x)	<1GB	
MATLAB/Python code	MATLAB/Jupyter notebook code	N	D	M/N	binary	<1GB	
Reports	presentation and discussion of results	N	D	T/N	.doc; .pdf	<1GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

For this project some open-source datasets will be used, and some data will be generated through simulation/experimentation. The URL of the open-source data are given here (url will be added/updated as per the use of open-source data)
https://engineering.case.edu/bearingdatacenter/download-data-file
https://phmsociety.org/public-data-sets/

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

N/A

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

N/A

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

N/A

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

An overview of experimental/simulation data will be maintained in the (Excel) files containing the Experimental results (e.g., tab sheet with list of source files interpreted).

· A dedicated **README.txt file** will attached in the folder of datasets
· All different binary files established with MATLAB/Python
· A short description will be added for each (group of) files listed.
· Reports are kept together with the above (in a shared storage place / folder).

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- Yes

RDR KU Leuven

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)

OneDrive (KU Leuven)
Shared network drive (J-drive)
High-volume data will be stored on Shared network drive J., and Low-volume data (such as, reports, PPTs, etc.) will be kept in OneDrive

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

OneDrive: sharing folders with only relevant persons
J-drive: password on folder

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

N/A

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- Shared network drive (J-drive)
- KU Leuven RDR

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

N/A

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as restricted data (upon approval, or institutional access only)

If access is restricted, please specify who will be able to access the data and under what conditions.

only members of our own research group.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- Other (specify below)

To be specified later.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

What are the expected costs for data sharing? How will these costs be covered?

N/A

Responsibilities

Who will manage data documentation and metadata during the research project?

Tauheed Mian (Postdoctoral fellow)
Pradeep Kundu (Supervisor)

Who will manage data storage and backup during the research project?

N/A (I.e., internal storage is used --> via ICTS KU Leuven)

Who will manage data preservation and sharing?

Pradeep Kundu
Tauheed Mian

Who will update and implement this DMP?

Tauheed Mian (Orcid ID: https://orcid.org/0000-0003-1332-7005)
Pradeep Kundu (Orcid ID: https://orcid.org/0000-0002-8336-5878)