Federated Learning PhD Research Proposal

Seán Ó Fithcheallaigh

# What is Federated Learning

Federated learning is a relatively new distributed machine learning paradigm that allows researchers to train models on large datasets while preserving the data's privacy and, of course, the source that generated the data, an essential requirement for many industries. Federated learning does this by providing a decentralised system that can be used to train neural networks on constrained devices (referred to as clients) using the data that the constrained device has gathered, where traditionally, the data would have been transmitted to a central server for training. When the client has trained a model on the stored data, the client can transmit those model updates back to the central server. When the server has gathered updates from enough clients, these updates can be brought together, and a global model can be trained, the model updates of which are then transmitted to the clients to use [1].

The fact that federated learning is an emerging field of study means that many open challenges can still be addressed during a research project.

# Problem Statement

Since the introduction of the Internet of Things (IoT), we live in a world where sensors have become commonplace, and the number of connected devices is projected to nearly double from today’s levels by 2030 [2]. Increasingly, sensors are being embedded in everyday objects (i.e., watches, cars, buildings, to devices to monitor a person's health), and these devices are generating massive amounts of data on a wide range of systems and devices.

As privacy concerns around data come increasingly into focus, federated learning is seen as a promising approach to allow the data gathered by a constrained device, or edge device, to be trained on that device. The model updates can then be sent back to the central server, where the updates from many edge devices can be brought together to train a new global model, which will be transmitted back to the edge devices. This gives federated learning the ability to keep the data it uses private and on the device itself, removing the need for the data to be transmitted back to a data centre for processing, as would be the case in more traditional machine learning processes.

As has been said above, federated learning is a relatively new area of study, and several challenges must be addressed. These include:

* Poor performance: Models can show reduced performance when dealing with heterogeneous data (i.e., non-IID) from various devices and environments [3].
* Communications: Improving communication protocols and overheads to allow communication between the client and the server in situations where bandwidth is limited, or connections are unreliable [3].
* Ineffective client operation: Clients can be ineffective for several reasons (unstable networks, energy constraints, reduced/limited computational capabilities), leading to poor system performance [4].
* Labelling: Labelling the client dataset can be challenging, as correct labelling is important in supervised learning tasks [5].
* Security: Federated learning systems are vulnerable to attacks on the data (data poisoning) or the model (model poisoning) [6].

The proposed research will address some of these challenges and add to the existing knowledge base in federated learning.

# Aims and Research Questions

The overarching aim of this research is to advance the understanding and application of federated learning wide ranging, distributed environments. The research will focus on improving and enhancing model efficiency, overall system security, and increasing communication efficiency.

To meet the research aim, a methodical process will need to be followed where one stage leads to the next, and so on. This process can be shown in the following research questions:

1. What is the current state of the art concerning federated learning:
   1. Identify the existing challenges, as well as the limitations, of federated learning.
   2. Detail this understanding in a comprehensive literature review.
2. From this understanding of the state of the art, identify several areas that require compressive research and analysis. For example:
   1. Improving model efficiency.
   2. Working with data heterogeneity.
   3. Improving communications schemes.
3. Develop a framework that is novel and addresses the current gaps or deficiencies in the state of the art:
   1. Test the framework against a clear set of criteria to rate performance.
   2. What improvements can be made to the framework.
   3. Can the framework be generalised for various data types or environments.
4. Evaluate the proposed framework using a series of real-world scenarios. While gathering empirical data, the framework should be evaluated for, among other things:
   1. Model performance.
   2. System Security.
   3. Data privacy.

# Project Plan

Please see Figure 1 below for the proposed project plan.

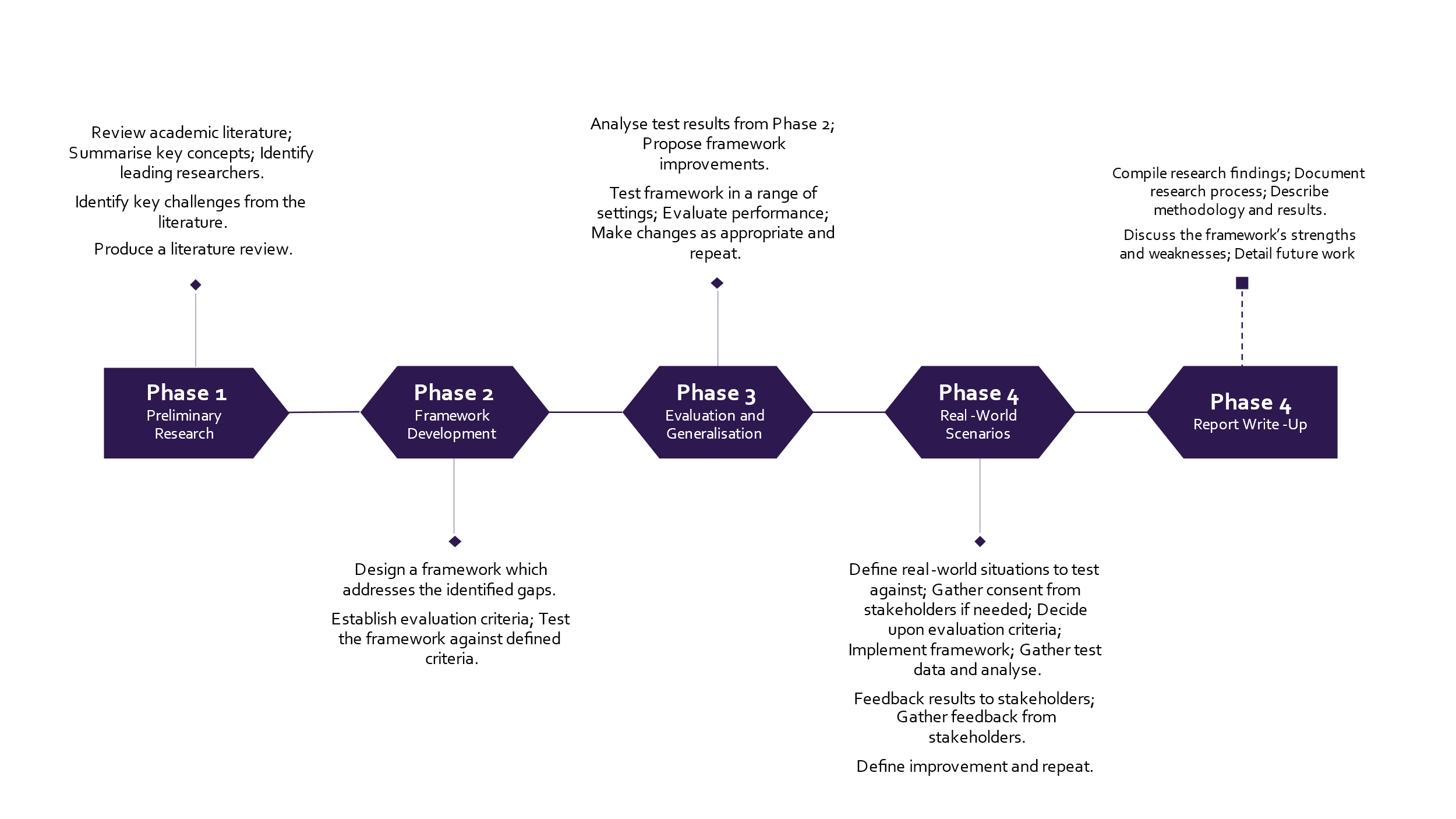


Figure : Proposed Project Plan

# References

|  |  |
| --- | --- |
| [1] | Hard, Andrew et. al., "Federated Learning for Mobile Keyboard Prediction," 2018. [Online]. Available: https://arxiv.org/abs/1811.03604. [Accessed 21 10 2023]. |
| [2] | L. Sujay Vailshery, "Number of Internet of Things (IoT) connected devices worldwide from 2019 to 2023, with forecasts from 2022 to 2030," Statista, July 2023. [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/. [Accessed 23 11 2]. |
| [3] | G. Pradeep Reddy and Y. V. Pavan Kumar, "A Beginner's Guide to Federated Learning," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*, Giza, Egypt, 2023. |
| [4] | F. Liu, M. Li, X. Liu, T. Xue, J. Ren and C. Zhang, "A Review of Federated Meta-Learning and Its Application in Cyberspace Security," *Electronics,* vol. 12, no. 15, p. 3295, 2023. |
| [5] | L. Li, Y. Fan, M. Tse and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering,* vol. 149, 2020. |
| [6] | S. Tyagi, I. Singh Rajput and R. Pandey, "Federated learning: Applications, Security hazards and Defense measures," in *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, Dehradun, India, 2023. |