# Human-AI Collaboration Cycle in The Development Stage of An AI-enabled System

**Pi-Yang Weng[1], Rua-Huan Tsaih[1], Hsin-Lu Chang[1]**

[1]Department of Management Information Systems, National Chengchi University (Taiwan)

**Presenter: Pi-Yang Weng**

Contact email: piyangong@gmail.com

# Pi-Yang Weng

**Pi-Yang Weng** has academic backgrounds in both Physics and Business Administration. He is currently a doctoral student interested in the research on Human-AI Collaboration and AI System Design at the Graduate School of Management Information System, National Chengchi University (Taiwan).

# Introduction

- Organization for Economic Cooperation and Development (OECD) has been promoting the concept of responsible AI with transparency, explainability, and accountability [1].

- Explainable AI (XAI) has garnered attention in the AI system development, especially in the high-stakes decision scenarios, such as medical and healthcare domains [2].

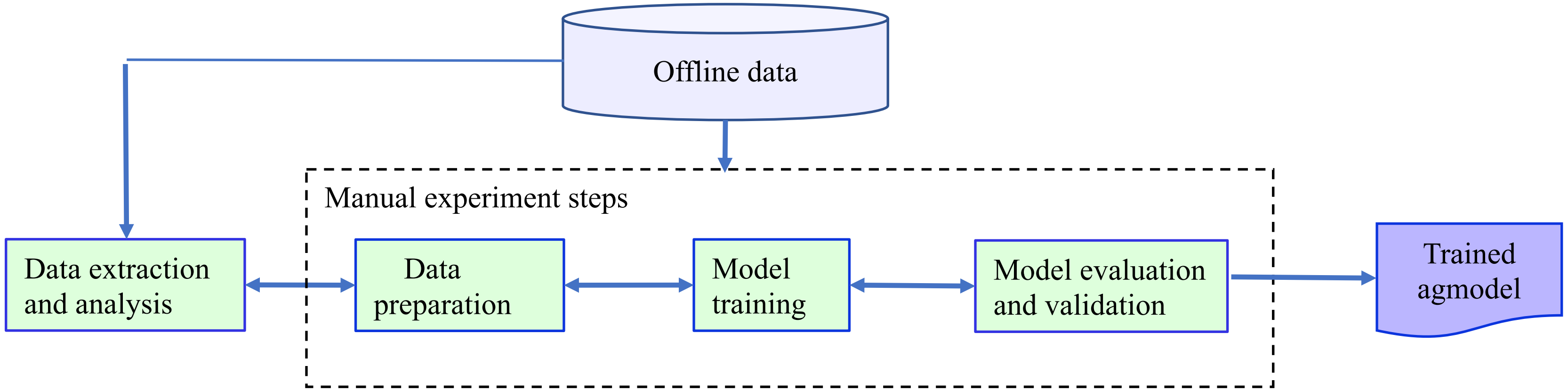
# Literature Review (1/3)

## Human-AI Collaboration

- Human and AI have different yet complementary capabilities [3].
- AI is not just a tool; it may become a teammate to enhance team performance [4].
- Human and AI can have mutual learning through which AI can learn from humans and humans can acquire insights from AI [5].
- ML needs methods that engage domain experts directly into the ML process and have them in the loop until the desired results are received [6].

# Literature Review (2/3)

## Machine Learning Pipeline




Source: Google Cloud Architecture Center [7]

The domain experts need to join the training data labeling task, in the case of supervised learning, for obtaining high-quality training datasets and avoiding garbage in, garbage out results [8].

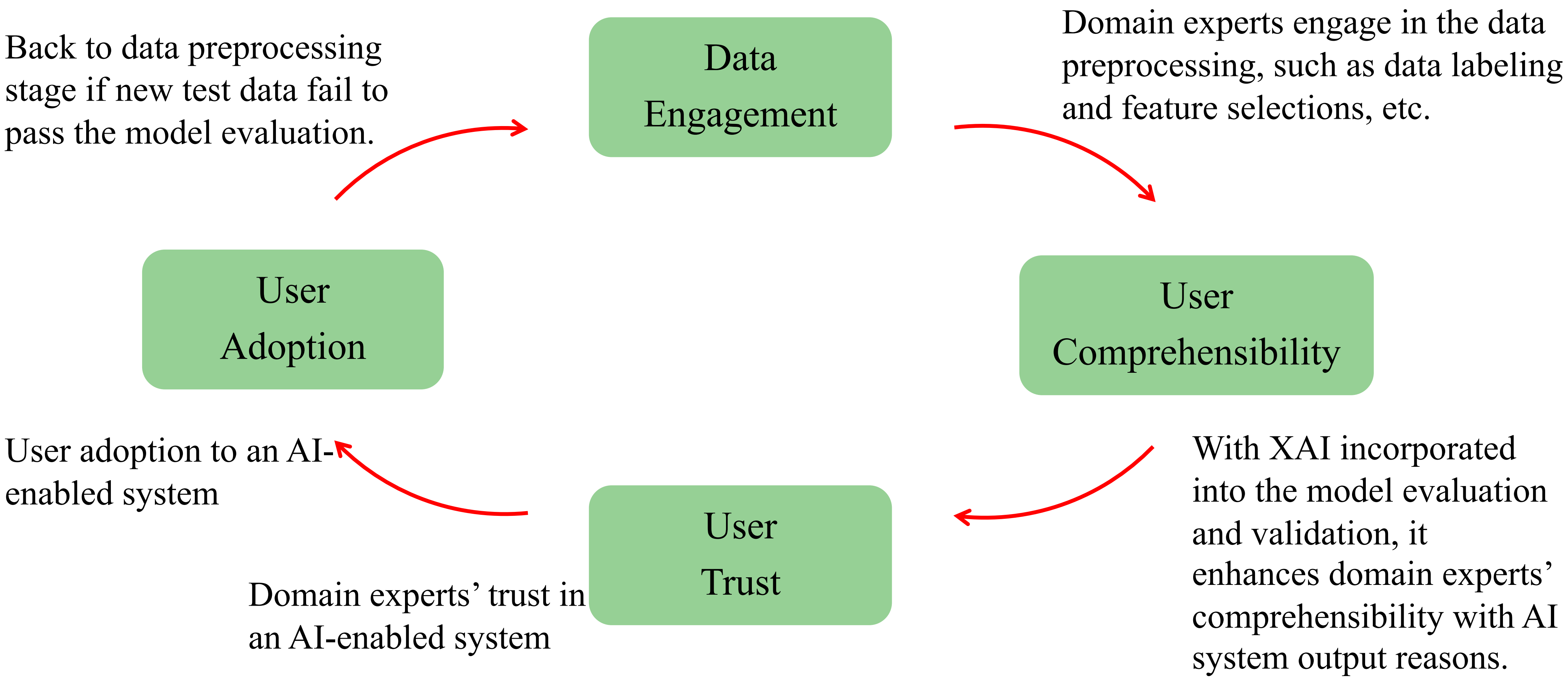
# Literature Review (3/3)

## Explainable AI (XAI)

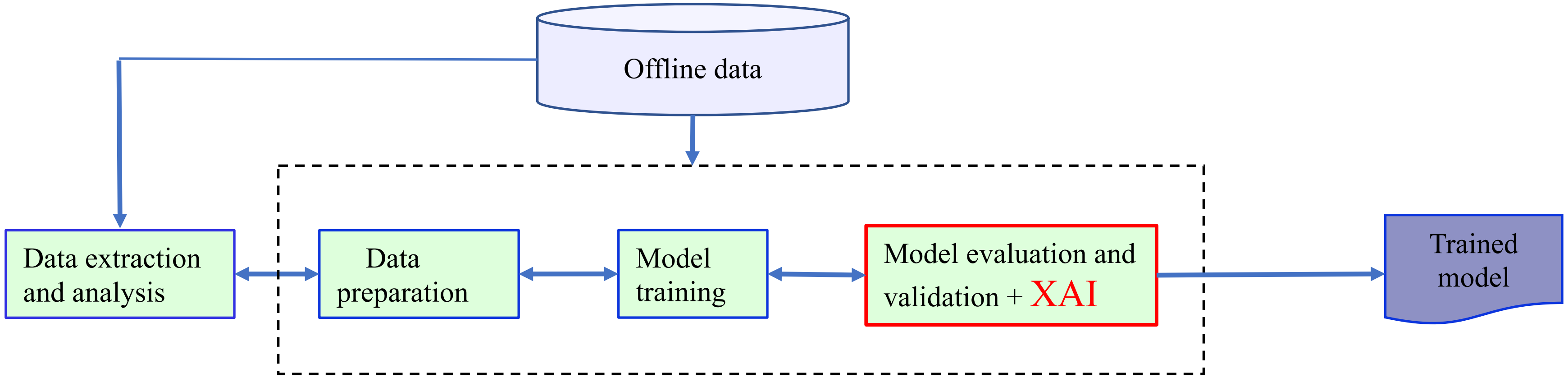
- XAI is a useful tool to unveil the ML black box and provides an explanation for each AI system output [9].

- XAI is especially instrumental in medicine and healthcare to ensure that the system outputs produced by the AI system are correct and justifiable [10].

# Human-AI Collaboration Cycle



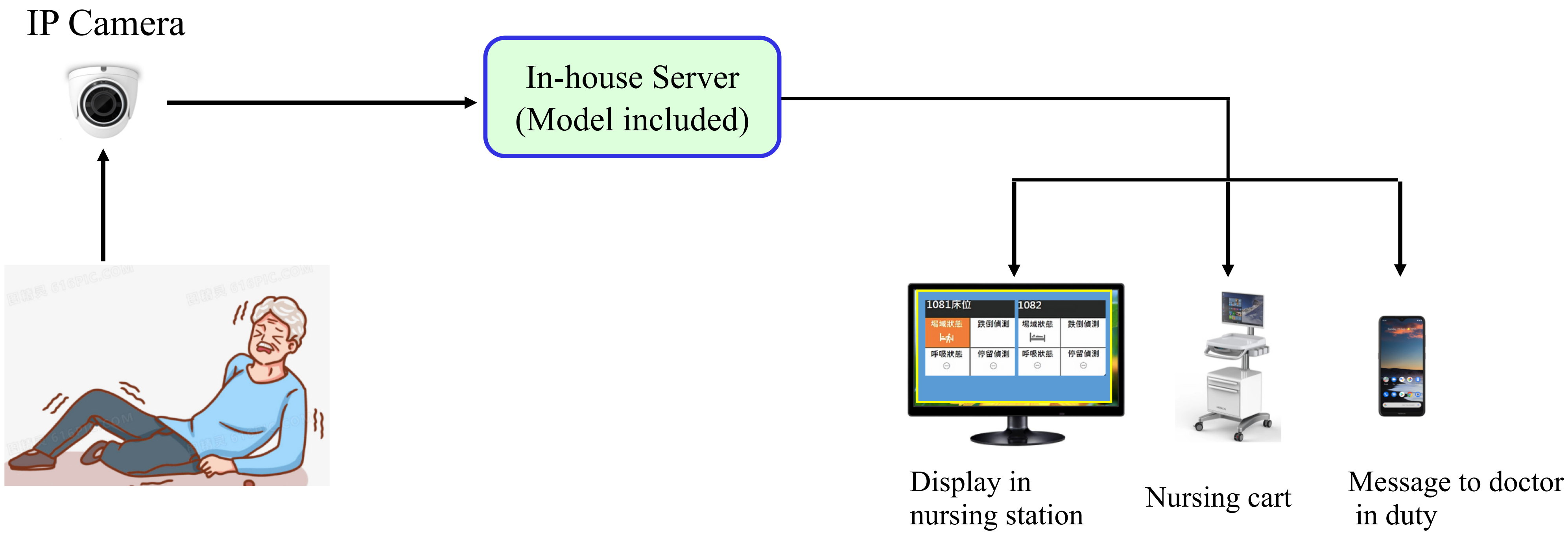
# ML Pipeline with XAI incorporated into model evaluation and validation process




**XAI is a useful tool to validate the model.**

# Research Methodology (1/3)

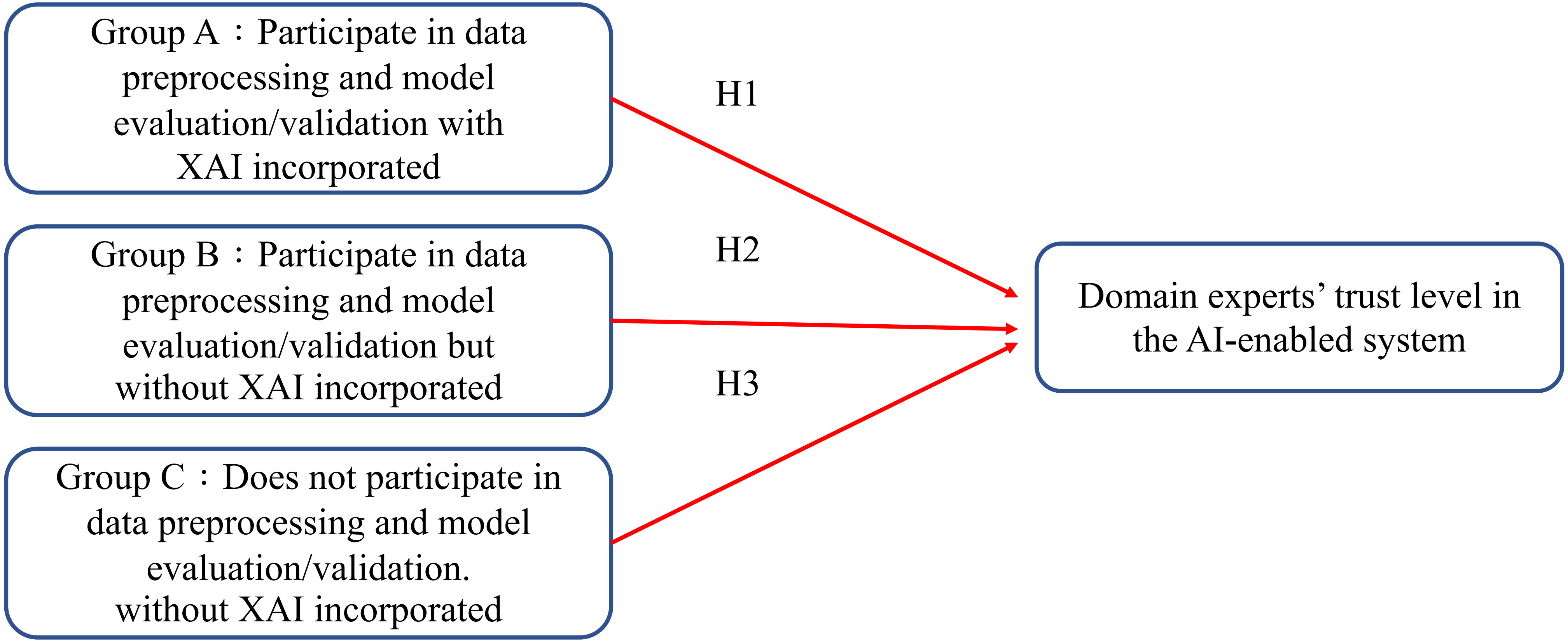
## IT Artifact




**AI-enabled Fall Detection System**

# Research Methodology (2/3)

## Hypotheses



# Research Methodology (3/3)

## Experiment Design

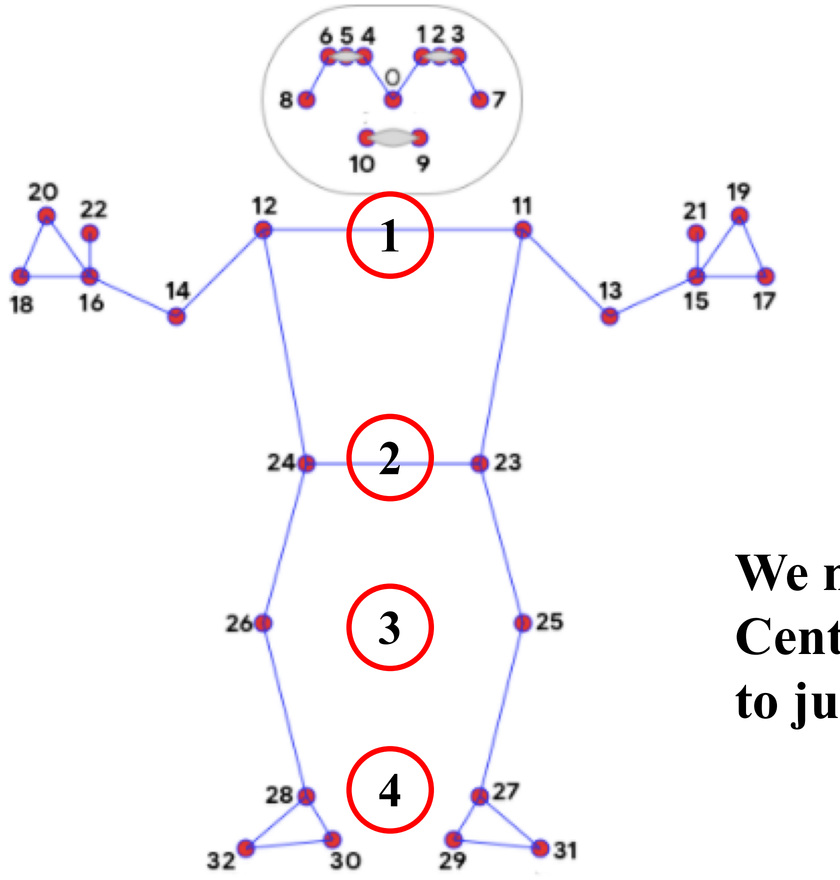
- Over 90 nurses from one local hospital will participate in this experiment.
- All nurses will be divided into three groups (Group A, B, and C)
- Each group will watch the same video demonstrating the brief introduction to the AI-enabled fall detection system.
- We designed different interaction modes with the AI-enabled system for each group:

  Group A: Participate in data preprocessing and data evaluation/validation with XAI incorporated.

  Group B: Participate in data preprocessing and data evaluation/validation but without XAI incorporated.

  Group C: Does not participate in data preprocessing and data evaluation/validation, also without XAI incorporated, to be informed of the system performance only. (As a control group)

# Human Skeleton

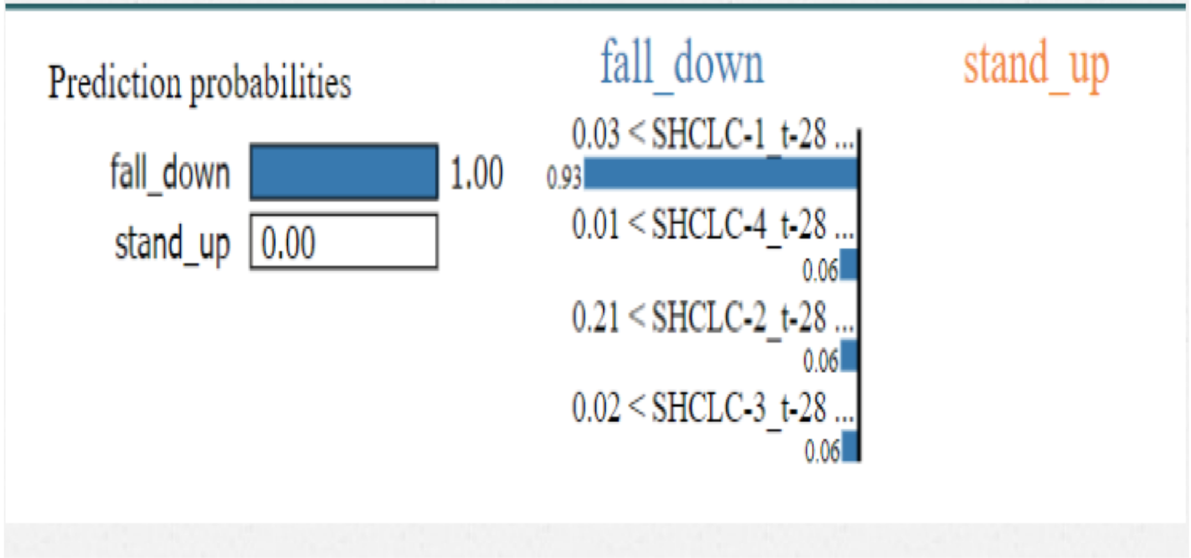



Point 1 to 4 represents the central line coordinate for shoulders, hips, knees, and ankles respectively.

**We measure the Speed of Human Central Line Coordinate (SHCLC) to judge a fall event.**

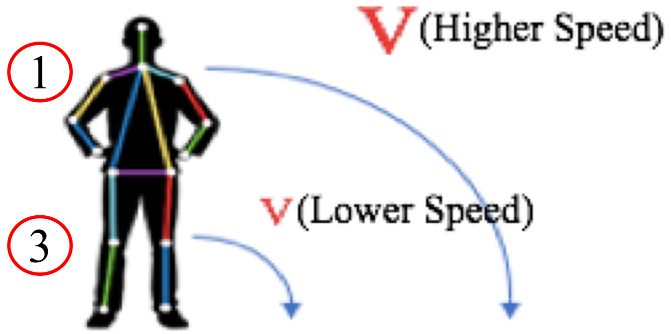
# LIME Output Format

## for an AI-enabled fall detection system



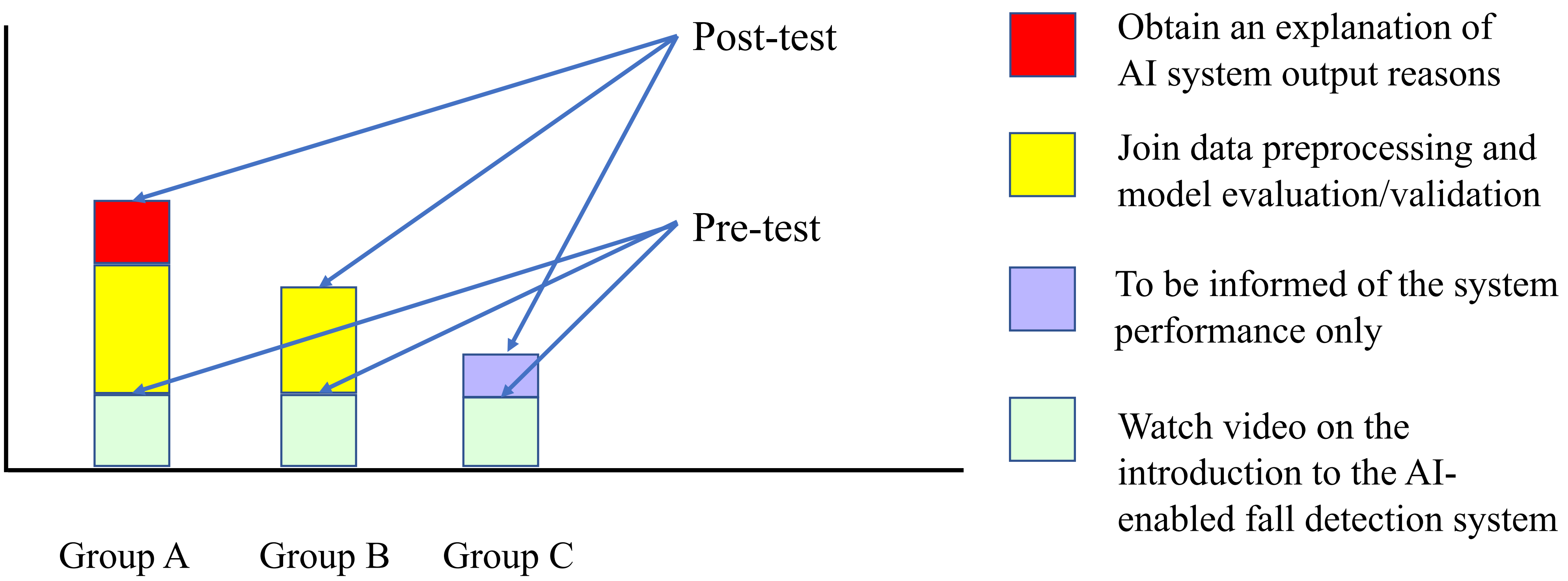

**This LIME output indicates that point 1 (SHCLC-1) has higher movement speed, which represents one kind of fall, e.g., fall over.**

# Different Movement Speed on Different Portion of a Human While Falling




**Point 1 has higher movement speed than point 3 while falling over.**

# The Timings of Pre-test and Post-test for Each Group



# Questionnaire

- I have confidence in the AI system performance.
- The AI system performance could be improved gradually.
- The output of the AI system is very predictable.
- The AI system is very reliable.
- The AI system is easy to use.
- The AI system is very efficient.
- The AI system can act as part of my team.
- I like to use the AI system.

(Modified from R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman (2018))

# Data Analysis Proposed

- ANOVA tool will be used for the significance analysis on trust level between groups.

- In addition to the quantitative analysis, we will observe the differences in their interaction modes with the AI-enabled system in each group and make a complete record for qualitative analysis.

- We have interest in the nurses' feedback or response to the XAI output explanation for one specific instance, which may encourage their data engagement.

# Research Contributions Expected

- This research will provide AI-enabled system designers with a HAC Cycle framework as a guideline for developing a responsible AI system.

- This research will highlight the importance of domain experts' engagement in the ML pipeline in the development stage of an AI-enabled system.

- This research will highlight the functionality of XAI incorporated in the model evaluation and validation process.

# References

[1] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL, 0449 G, 2022. https://www.oecd.org/digital/artificial-intelligence [retrieved: February, 2024].

[2] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, Survey of explainable AI techniques in healthcare. Sensors, 23, 634, pp. 1-19, 2023.

[3] M. H. Jarrahi, Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Kelly School of Business, Indiana University, BUSHOR-1478, pp. 1-10, 2018.

[4] I, Seeber et al., Machines as teammates: A research agenda on AI in team collaboration. Information and Management, Vol. 57, pp. 1-22, 2020.

[5] M. J. saenz, E. R. Revilla, and C. Simon, Designing AI systems with human-machine teams. MIT Sloan Management, Reprint# 61430, 2020.

[6] M. Maadi, H. A. Khorshidi, and U. Ackelin, A review on human-AI interaction in machine learning and insights for medical applications.

[7] Google Cloud, MLOps level 0: Manual process, Google Cloud Architecture Center, 2020. https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines [retrieved: February, 2024].

[8] R. S. Geiger et al., “Garbage in, garbage out” revisited: What do machine learning application papers report about human-labeled training data ? Quantitative Science Studies, Vol. 2, No. 2, pp. 1-42, 2021.

[9] M. T. Rebeiro, S. Signh, and C. “Guestrin, Why should I trust you” Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135-1144, 2016.

[10] A. Shaban-Nejad, M. Michalowski, and D. L. Buckeridge, Explainability and interpretability: Keys to keep medicine. Explainable AI in Healthcare and Medicine, Vol. 914, pp. 1-10, 2021.

[11] R. R. Hoffman, S. T. Mueller, and G. Klein, and J. Litman, Measuring trust in the XAI context. Technical Report, DARPA Explainable AI Program, pp. 1-26, 2018.


Thank You!