

Towards Minimally Domain-Dependent and Privacy-Preserving Architecture and Algorithms for Digital Me Services: EdNet and MIMIC-III Experiments

Kyoung Jun Lee¹, Baek Jeong¹, Youngchan Kim², Suhyeon Kim¹

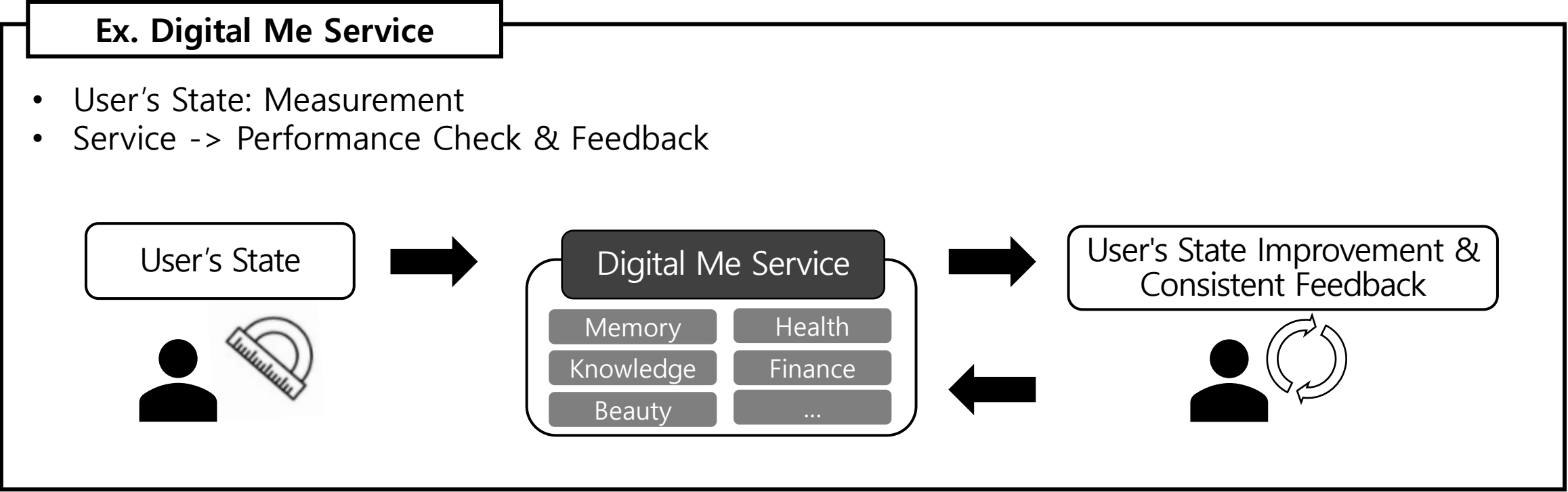
¹ Department of Big Data Analytics, Kyung Hee University, Korea

² Graduate School of Artificial Intelligence, POSTECH, Korea

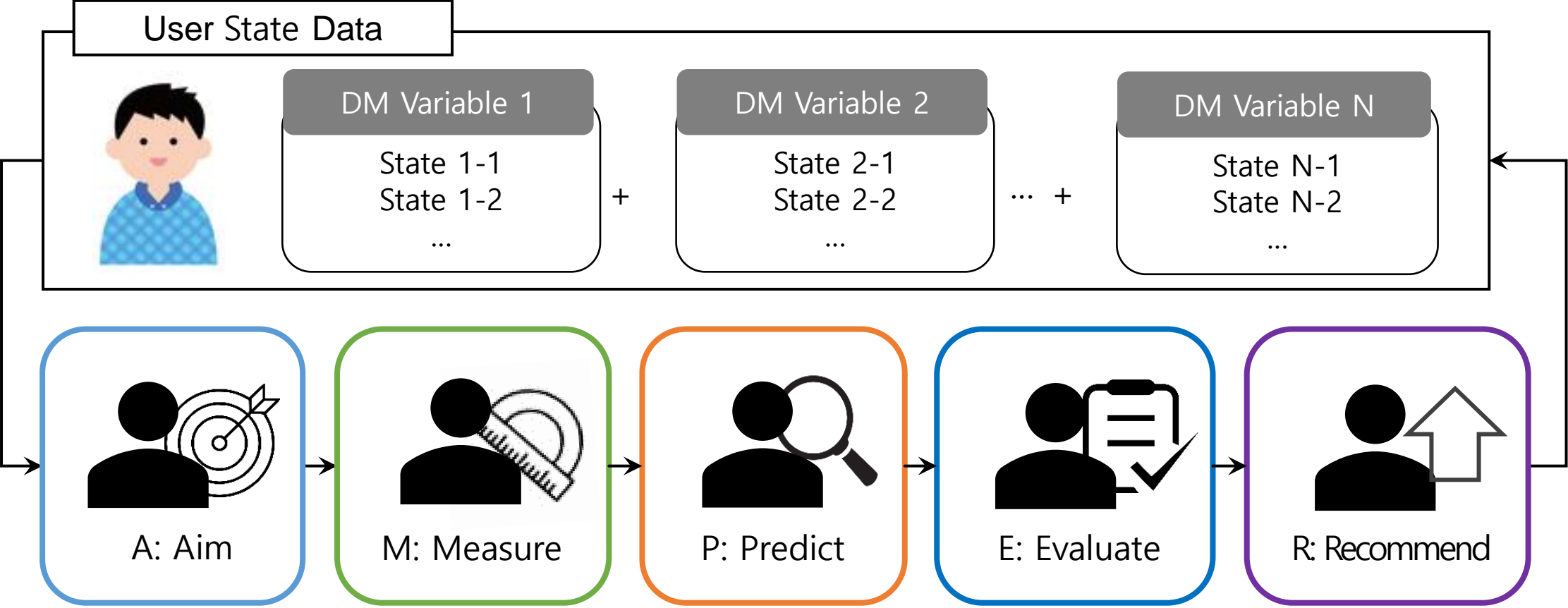
Digital Me Service

Introducing an AI-based product service system that **makes it possible to manage the individual's state** (health, beauty, memory, knowledge, finance, happiness, etc.) in real-time.

- The Digital Me service provides a personalized experience while ensuring the security and privacy of user data. It accurately measures the current state of users and predicts future states to recommend optimal actions.



AMPER(Aim-Measure-Predict-Evaluate-Recommend)



Minimally Domain-Dependent Digital Me algorithms

Based on the user state data, in Digital Me algorithmic AMPER structures that enhance the user's state, develop models by accessing behavior data in a general method without specifying which domain it utilizes.

In user state data, approached in terms of Digital Me variable and value of user behavior according to each variable.

- DM Variable: Question to measure the English proficiency (Education), Heart rate, blood pressure, etc. to measure health (Health)
- DM Value/Behavior: Time spent solving question, user answers (Education), systolic and diastolic blood pressure levels (Health)

Using Riid's EdNet Data
as general method

| DM Variable Behavior1 Behavior2 | | | | Next Behavior | | | |
|---------------------------------|-------|-----|-------|---------------|---------------------------|--|--|
| 0 | q6952 | d | 23250 | 0 | <esp> q6952 d 23250 <esp> | | |
| 1 | q6949 | b | 2750 | 1 | <esp> q6949 b 2750 <esp> | | |
| 2 | q6952 | a | 49250 | 2 | <esp> q6952 a 49250 <esp> | | |
| 3 | q490 | b | 61000 | 3 | <esp> q490 b 61000 <esp> | | |
| 4 | q842 | a | 9000 | 4 | <esp> q842 a 9000 <esp> | | |
| ... | ... | ... | ... | ... | ... | | |
| 29796 | q4070 | c | 25000 | 29796 | <esp> q4070 c 25000 <esp> | | |

Exploration with EdNet Data

The EdNet dataset includes the timestamp (indicating when a user began answering a specific question), question ID, the user's response, and elapsed time (duration taken to answer the question).

| user | timestamp | question_id | part | tags | correct_answer | user_answer | elapsed_time |
|---------|-----------|-------------|------|----------------|----------------|-------------|--------------|
| u100240 | 1.516E+12 | q176 | 1 | 6;7;183 | d | d | 30000 |
| u100240 | 1.516E+12 | q1279 | 2 | 24;26;182;184 | c | c | 15000 |
| u100240 | 1.516E+12 | q2067 | 3 | 52;183;184 | b | b | 41666 |
| u100240 | 1.516E+12 | q2068 | 3 | 55;183;184 | a | b | 41666 |
| u100240 | 1.516E+12 | q2069 | 3 | 179;52;183;184 | d | d | 41666 |
| u100240 | 1.516E+12 | q3412 | 4 | 64;52;184 | a | a | 23666 |
| u100240 | 1.516E+12 | q3413 | 4 | 64;52;184 | d | b | 23666 |
| u100240 | 1.516E+12 | q3411 | 4 | 64;53;184 | c | a | 23666 |
| u100240 | 1.516E+12 | q2991 | 4 | 59;52;183 | c | c | 19333 |
| u100240 | 1.516E+12 | q2993 | 4 | 59;52;183 | b | c | 19333 |
| ... | ... | ... | ... | ... | ... | ... | ... |

Exploration with MIMIC-III Data

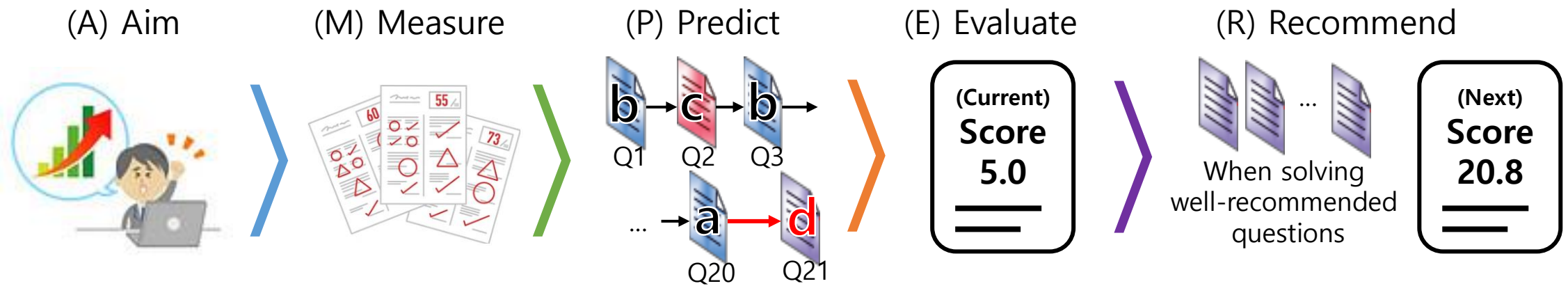
The Medical Information Mart for Intensive Care III (MIMIC-III) database, developed through a collaboration with MIT, is sourced from the intensive care units of Beth Israel Deaconess Medical Center.

- It includes de-identified health records for 61,532 patients admitted to intensive care between June 2001 and October 2012, comprising data for 53,432 adults and 8,100 infants. The dataset includes patient demographics, vital signs, laboratory results, medication details, caregiver notes, imaging data, and mortality statuses.
 - Specifically, prescription data exists for treatments to learn the type of medication and dosage form of the treatment.
- Prescription Data Example

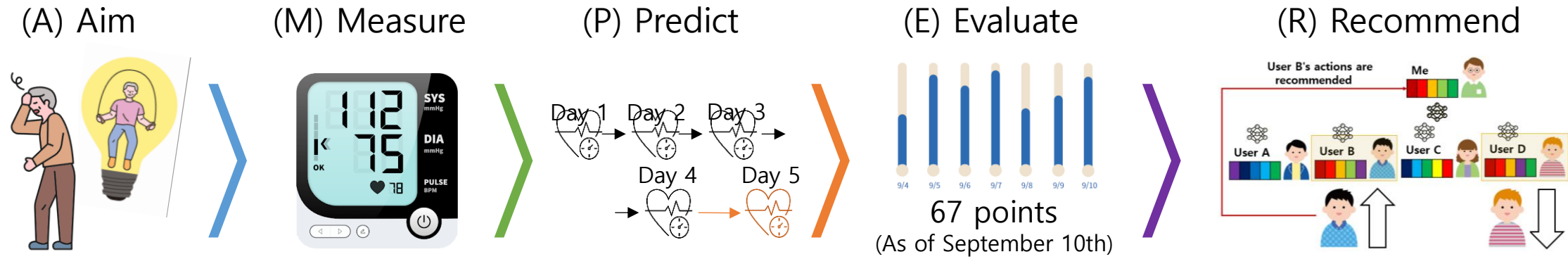
| row_id | subject_id | hadm_id | icustay_id | startdate | enddate | drug_type | drug | formulary_drug_cd | gsn | prod_strength | route | ... |
|--------|------------|---------|------------|-----------------|-----------------|-----------|----------------------|-------------------|-------|---------------------------|-------|-----|
| 86575 | 43735 | 112662 | 293363 | 2128-11-05 0:00 | 2128-11-09 0:00 | MAIN | Pantoprazole | PANT40 | 27462 | 40mg Tablet | PO | ... |
| 86576 | 43735 | 112662 | 293363 | 2128-11-05 0:00 | 2128-11-06 0:00 | MAIN | Furosemide | FURO40I | 8205 | 40mg/4mL Vial | IV | ... |
| 86577 | 43735 | 112662 | 293363 | 2128-11-05 0:00 | 2128-11-05 0:00 | MAIN | Carvedilol | CARV3125 | 28108 | 3.125mg Tablet | PO | ... |
| 86578 | 43735 | 112662 | 293363 | 2128-11-05 0:00 | 2128-11-05 0:00 | MAIN | Carvedilol | CARV3125 | 28108 | 3.125mg Tablet | PO | ... |
| 86579 | 43735 | 112662 | 293363 | 2128-11-05 0:00 | 2128-11-09 0:00 | MAIN | Metoprolol Tartrate | METO25 | 50631 | 25mg Tablet | PO | ... |
| 86580 | 43735 | 112662 | 293363 | 2128-11-06 0:00 | 2128-11-09 0:00 | MAIN | Heparin | HEPA5I | 6549 | 5000 Units / mL- 1mL Vial | SC | ... |
| 86605 | 43735 | 112662 | 293363 | 2128-11-06 0:00 | 2128-11-06 0:00 | BASE | 0.9% Sodium Chloride | NS500 | 1210 | 500mL Bag | IV | ... |
| 86606 | 43735 | 112662 | | 2128-11-08 0:00 | 2128-11-09 0:00 | BASE | Iso-Osmotic Dextrose | VANCOBASE | | 200ml Bag | IV | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

AMPER Experimental Results Using EdNet and MIMIC-III

(Education) Improving English skills

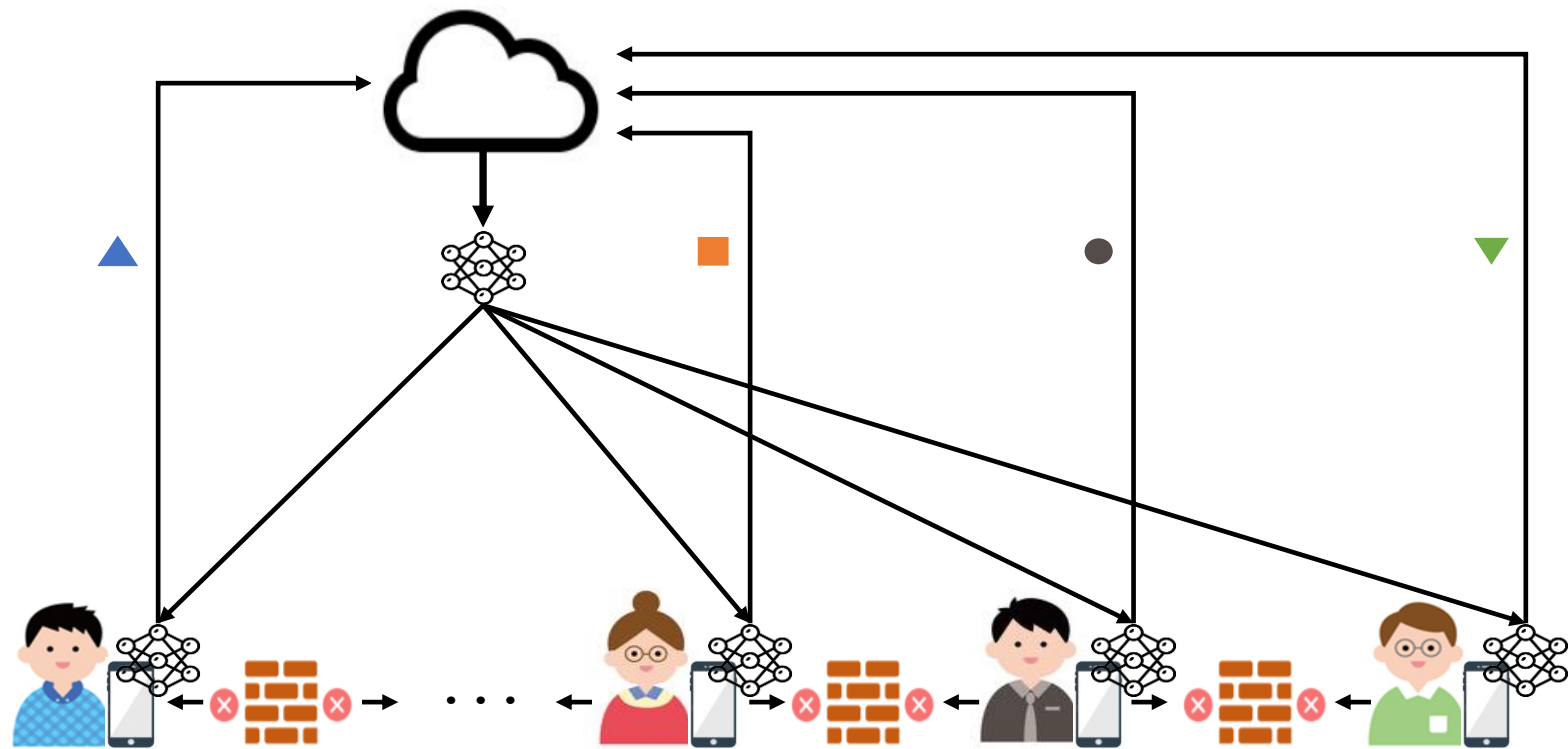


(Healthcare) Blood pressure management



Personalized Federated Learning Experiment

Personalized Federated Learning (PFL) is described as a method inspired by Model-agnostic Meta Learning that identifies robust initializations to quickly adapt to varying data distributions of individual devices, ensuring privacy by storing data locally while also enhancing prediction performance.

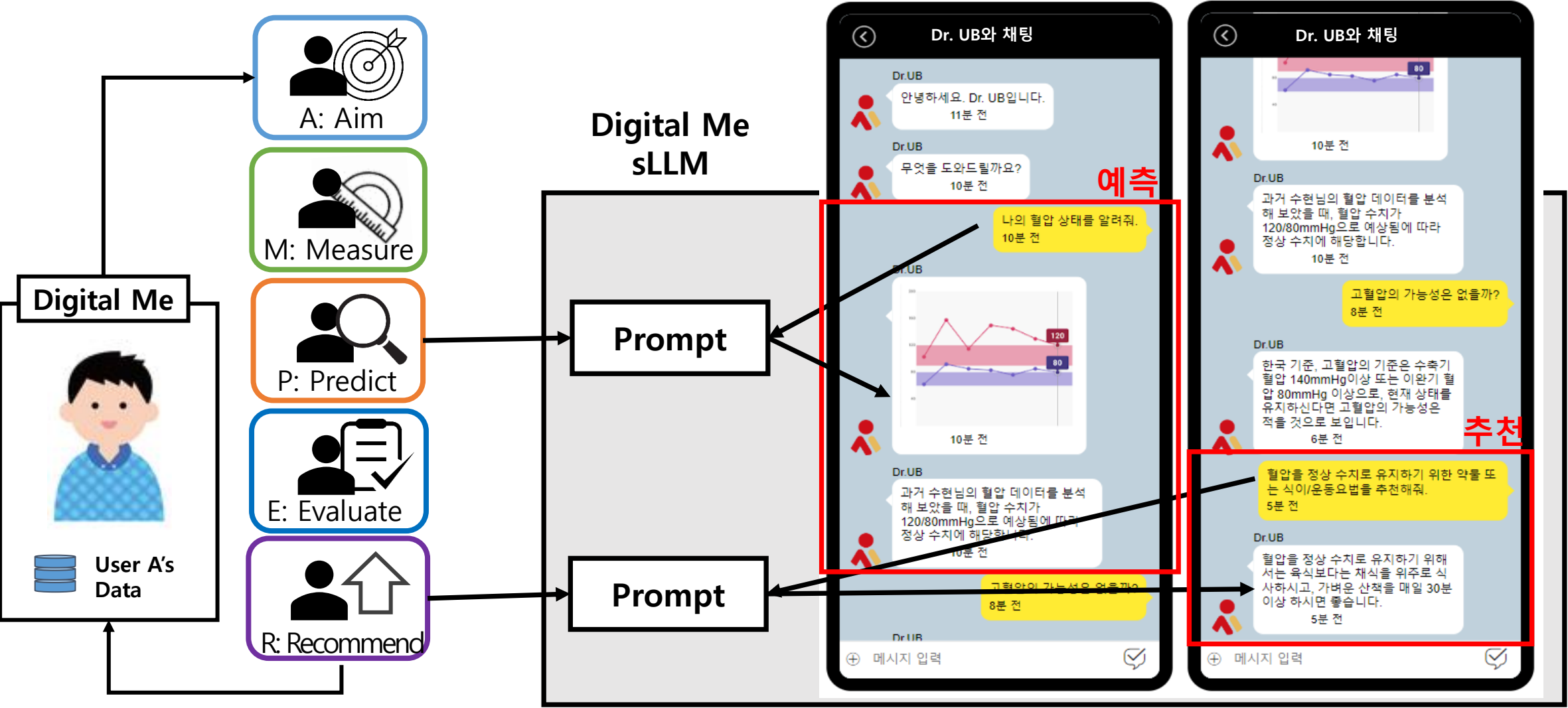


(P) Predict on MIMIC-III Data with PFL

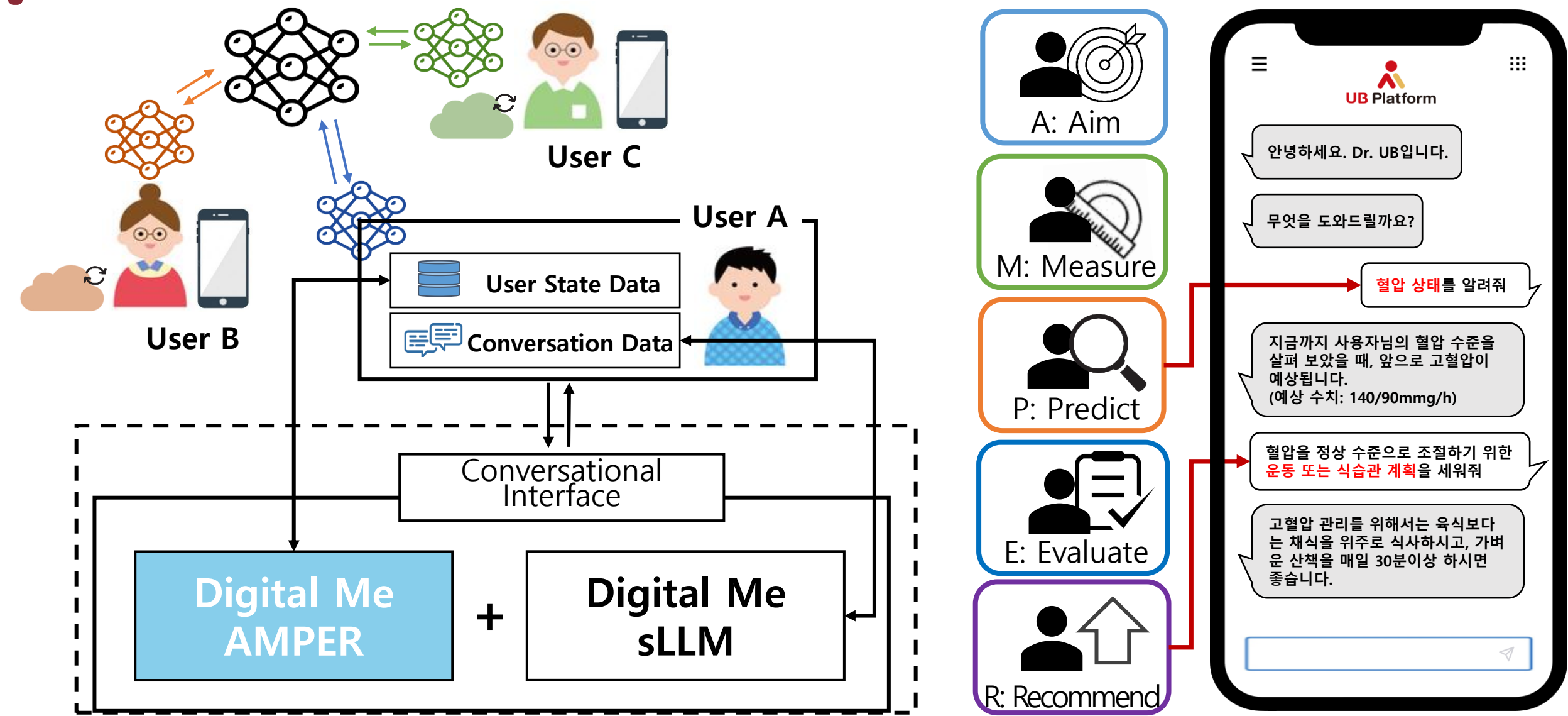
- Grouped by patient number and distributed to individual (virtual) devices
- Total 9,908 patients / 2,065,091 cases of data (approximately 208 days of data per person)
 - Max: 26,282 days of data
 - Min: 2 days of data retention

| | Personalized Federated Learning(PFL) | | Data Sharing |
|-----|--------------------------------------|------------------------------|--|
| | Data from a total of 9,908 people | | |
| | Local Model | PFL Model (Global_weight) | 2,065,091 total data (9,908 people) |
| MAE | 11.70mmHg | 8.03mmHg | 7.15mmHg |

Conversational Digital Me Service Interface



Privacy-Preserving Digital Me with PFL and s-LLM



Conclusion and Outlook

Minimally Domain-Dependent Importance

- This study demonstrates the practical feasibility of applying Digital Me services across diverse domains, including education, healthcare, personal financial management, and lifestyle coaching.

Enhanced Privacy with PFL and s-LLM

- Personalized Federated Learning (PFL) processes data locally on devices to significantly enhance privacy. Additionally, small Language Models (s-LLMs) independently manage sensitive personal details on devices, safeguarding user privacy.

Real-World Challenges and Scalability

- Addressing scalability and privacy-performance balance is crucial in diverse environments. PFL faces challenges in managing data heterogeneity, impacting model convergence and effectiveness.