# How to Handle Wellbeing in Socially Responsible AI? - Findings from Sleep Perspective -

Keiki Takadama [1]

[1] *The University of Electro-Communications, 1-5-1, Chofugaoka, Chofu, Tokyo 182-8585 Japan*

**Abstract**

This paper focuses on Socially Responsible AI (SRAI) and discusses how SRAI should handle wellbeing. From the viewpoint of the AI ethics in SRAI, this paper claims that the seven issues of the AI ethics are not enough for the wellbeing systems such as a healthcare system, *i.e.*, the "adaptability" is needed as the new concept to cope with wellbeing because health condition always changes. To investigate this issue, this paper focuses on sleep and conducts the human subject experiment on the sleep stage estimation and has reveal the following implications: (1) own data (*i.e.*, heartrate in this experiment) contributes to improving the accuracy of the sleep stage estimation in many cases but it is not useful in health condition change; (2) others' data with the highest similarity of the target person is useful even in health condition change, which suggests that others' data contributes to providing the "adaptability" in the sleep monitoring systems.

**Keywords**

socially responsible AI, adaptability, wellbeing, sleep stage estimation

## 1. Introduction

As AI technologies grow, we receive benefits of them, but may take a risk of using AI systems at the same time (such as privacy information leakage). For such risks, one question comes up, that is, who should take the responsibility for it? Since AI cannot take responsibility to what AI did, companies that developed the systems may have the responsibility (in the case of system malfunction) or users that used the systems may have the responsibility (in the case of not-proper use). This is the fundamental problem when using AI systems. For this issue, socially responsible AI (SRAI) [2] has been recently attracted much attention on to address the responsibility of AI.

In detail, SRAI is addressed by the technology and ethic viewpoints. From the technology viewpoint, the trustworthy AI plays an important role, and explainable AI [1] and human centered design are the important technologies. From the ethic viewpoint, on the other hand, the following issues are taken account of as the AI ethics: accountability, transparency, fairness, reliability/ safety, security, inclusiveness, cooperativity (see the next section for details). In particular, these issues in the AI ethics are very important for companies to provide reliable products, but they may not be enough from the viewpoint of wellbeing (such as healthcare systems). This is because health condition always changes, *i.e.*, correct answer/suggestions changes according to its condition change. From the data viewpoint, currently provided data is not always correct, *e.g.*, today's data may not be useful for tomorrow.

To address this issue, this paper proposes to add the new concept in SRAI, that is, "adaptability" to develop the system that can adapt to changes of human conditions or environments surrounding us. To investigate how SRAI should handle well-being, this paper focuses on the sleep as one of targets of the healthcare systems and conducts the human subject experiment on the sleep stage estimation.

This paper organized as follows. The next section briefly explains socially responsible AI, and Section 3 describes the real-time sleep stage estimation. The human subject experiment is conducted in Section 4. Finally, our conclusion is given in Section 5.

AAAI 2023 Spring Symposia, Socially Responsible AI for Well-being, March 27–29, 2022, USA
EMAIL: keiki@inf.uec.ac.jp (A. 1);

## 2. Socially responsible AI (SRAI)

The AI ethics in SRAI addresses the following seven issues.

- **Accountability:** How AI system works should be explained.
- **Transparency:** Mechanisms of AI system should be public for all persons (such as open source).
- **Fairness:** Data should not be biased (such as race and gender data).
- **Reliability/safety:** AI system should provide sure and/or safe systems.
- **Security:** Privacy information or data should be protected.
- **Inclusiveness:** All persons should receive benefits of AI system.
- **Cooperativity:** AI system should support and/or cooperate with people.

From the viewpoint of companies, Microsoft addresses the above issues except for cooperativity [5]. Google addresses the above issues except for transparency and cooperativity but adds the following issues [3]: (i) uphold high standards of scientific excellence, and (ii) be made available for uses that accord with these principles.

## 3. Realtime sleep stage estimation

The real-time sleep stage estimation (RSSE) method [4] estimates the sleep stage from the heartrate acquired from a mattress sensor without connecting any devices and/or electrodes to the human body.

### 3.1. Mechanism

To estimate the sleep stage, the RSSE method generates a model of the middle frequency component of the heartrate during sleep as the regression of the trigonometric function. Using this model, the RSSE method predicts prospective heartrate from the partially obtained heartrate, and estimates the sleep stage in real-time. Concretely, the middle frequency component of the heartrate is modeled as follows,

$$h(t,\phi) = c + \sum_{n=1}^{N} \left( a_n cos\left(\frac{2\pi t}{L/n}\right) + b_n sin\left(\frac{2\pi t}{L/n}\right) \right), \quad (1)$$

where $h(t, \varphi)$ denotes the predicted heartrate at time $t$ with the model parameter $\varphi = \{c, a_1, \dots, a_N, b_1, \dots, b_N\}$, $L$ denotes the maximum period of the middle frequency component, and $N$ denotes the number of composed trigonometric functions. The model parameters $\varphi$ are provided by the maximum likelihood estimation method from the following likelihood formula,

$$J(\phi) = \frac{1}{T}\sum_{t=1}^{T} \left(HR(t) - h(t,\phi)\right)^2 + \lambda P(\phi), \quad (2)$$

where $T$ denotes the elapsed time after falling asleep, and $HR(t)$ denotes the obtained heartrate at time $t$. In detail, the first term calculates the mean square error between the estimated and the obtained heartrate for each time, while the second term, $P(\varphi)$, denotes the penalty function for the model parameter. $\lambda$ has a role to balance between the mean square error and the penalty function. As the penalty function $P(\varphi)$, the following equation is employed:

$$P(\phi) = \frac{1}{N}\sum_{n=1}^{N} \left(a_n^2 + b_n^2\right), \quad (3)$$

which is the normalized term that penalizes the large parameter values to avoid over-fitting to training data. Such technique is usually used in a machine learning algorithm. After calculating the parameters $\varphi$, the prospective heartrate can be predicted, and the sleep stage is estimated by discretizing the predicted heart rate $h(t, \varphi)$ according to the following formula:

$$s(t) = \begin{cases} 5 & \left\lceil \frac{h(t,\phi)-ave}{stdev} + 2 \right\rceil > 5 \\ 0 & \left\lceil \frac{h(t,\phi)-ave}{stdev} + 2 \right\rceil < 0 \\ \left\lceil \frac{h(t,\phi)-ave}{stdev} + 2 \right\rceil & otherwise \end{cases} \quad (4)$$

where $s(t)$ denotes the sleep stage at time $t$, while $ave$ and $stdev$ denote the average and the standard deviation of the predicted heartrate $h(t, \varphi)$, respectively. $\lceil x \rceil$ denotes the ceiling function that returns the minimum integer value which is equal to or greater than $x$. After discretization, the value from 5 to 0 is assigned to the sleep stages of WAKE, REM, Non-REM1, Non-REM2, Non-REM3, and Non-REM4, respectively.

### 3.2. Algorithm

The detailed algorithm of the RSSE method is described in Algorithm 1. After detecting falling asleep, the heart rate $HR(t_{now})$ at time $t_{now}$ is measured, and the model parameters are calculated to minimize Eq. (2) when the predefined estimation interval term $t_{int}$ has passed. Using these parameters, the prospective heartrate is predicted and the current sleep stage is calculated by discretizing the predicted heartrate according to Eq. (4). After estimating current sleep stage, these processes are repeated until awake.

**Algorithm 1** The RSSE method

$t_{prev} = 0$
**while** sleep **do**
  $t_{now} = current\_time$
  Measure heart rate $HR(t_{now})$ at time $t_{now}$
  **if** $t_{now} - t_{prev} \geq t_{int}$ **then**
    Calculate parameters $\phi$ that minimizes $J(\phi)$ in equation (2)
    Estimate entire heart rate $h(t, \phi)$ from equation (1) according to the calculated parameters for $t = [0, max(T, L)]$
    Estimate sleep stage from equation (4) according to estimated $h(t, \phi)$
    Output current sleep stage from $t_{prev}$ to $t_{now}$
    $t_{prev} = t_{now}$
  **end if**
**end while**

## 3.3. Difficulty of Sleep stage estimation in health condition change

As described in the previous subsections, the RSSE method estimates the sleep stage by calculating the middle frequency component of the heartrate. However, its accuracy is affected by the initial model parameter $\varphi = \{c, a_1, \dots, a_N, b_1, \dots, b_N\}$, *i.e.*, some initial parameters may overfit to the given heartrate. For this issue, the RSSE employs the model parameter $\varphi$ of the same person calculated in the past day as the initial value. This is because a tendency of sleep is generally similar in the same person, which means that a tendency of the middle frequency component of the heartrate is also similar in the same person.

What should be noted here, however, is that this is not true when his/her health condition changes. In such case, the model parameter $\varphi$ of "the other person" may be useful. Figure 1 shows the example of the above cases. When focusing on the same person, the tendency of heartrate on 3/27 and 3/28 are different from that on 3/29 even in the same person. This means that the past own data (3/27 and 3/28) may not be useful for the today's estimation (3/29). When focusing on the other persons, on the other hand, the tendency of the heartrate of the left person on 2/9 is different from that of the middle person on 3/15 but is similar to that of the target person on 3/29 even in the different persons. This means that the other's data (3/15) may be useful for the target person's estimation (3/29). For this issue, our previous method employs the model parameter $\varphi$ of his/her own past heartrate or others' heartrate which has the highest similarity of the heartrate of the target person [7].

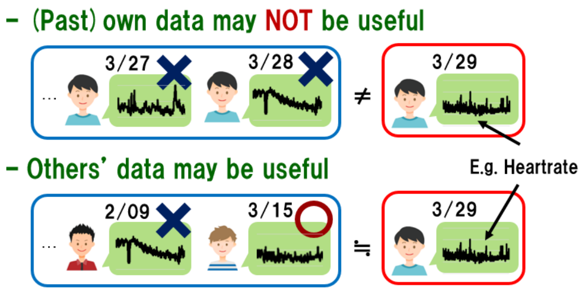


**Figure 1**: Similarity of heartrate with own data and others' data

## 4. Human subject experiment

### 4.1. Experimental design

To investigate how the sleep stage estimation is affected by own data and others' data, this paper conducts the human subject experiment. Concretely, the following two cases are investigated:

- **Case 1:** Sleep stage estimation with/without own data (*i.e.*, the initial model parameter of "the same person" calculated in the past day)
- **Case 2:** Sleep stage estimation with/without others' data (*i.e.*, the initial model parameter of "the other person")

Note that the same person of 20s, 40s, and 60s are employed in case 1, while the different persons of 20s, 30s, 40s, 50s, and 60s are employed in case 2.

As evaluation criterion, this paper employs an accuracy of the sleep stage estimation (of 6 stages) compared with the polysomnography (PSG) test as the gold standard method. In the PSG test, biological data of electroencephalography (EEG), electrooculogram (EOG), and electromyogram (EMG) are acquired to determine the sleep stage by the R&K method [6].

### 4.2. Experimental result

Figure 2 shows the accuracy of the sleep stage estimation with/without own data, where the vertical and horizontal axes indicate the accuracy of the sleep stage estimation and 3 days of the three human subjects, respectively. In detail, the blue and red bars indicate the accuracy without and with own data, respectively. The marks "M" and "F" indicate male and female, respectively.

The number in the parentheses indicate the age (*e.g.*, (20) indicates the person of 20s). From this figure, the accuracy of the sleep stage estimation with own data is higher than that without own data except for the day 1 in the male person of 20s, the day 2 in the male person of 40s, and the day 2 in the female person of 60s, which are days of bad health condition. This result suggests that own data is useful to increase the accuracy of the sleep stage estimation in many days but not useful in bad health condition.

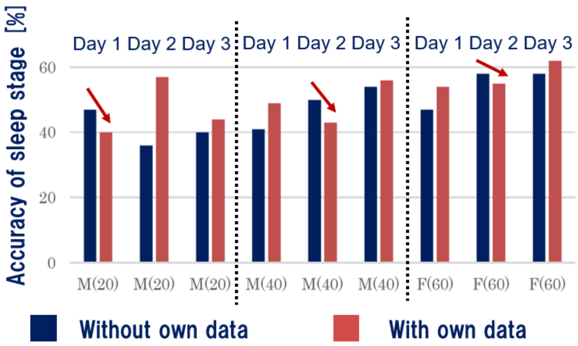


**Figure 2**: Accuracy of sleep stage estimation with/without own data

Compare with Figure 2, Figure 3 shows the accuracy of the sleep stage estimation with/without others' data, where the vertical axis has the same meaning of Figure 2 while the horizontal axis indicates 1 day of the six human subjects. The blue and red bars indicate the accuracy without and with others' data, which have the highest similarity of the heartrate of the target person, respectively. The marks "M" and "F", and the number in the parentheses have the same meaning of Figure 2. From this figure, the accuracy of the sleep stage estimation with others' data is the same or higher than that without others' data in all persons. Considering the fact that M(20), M(40), are F(60) in the red curved square

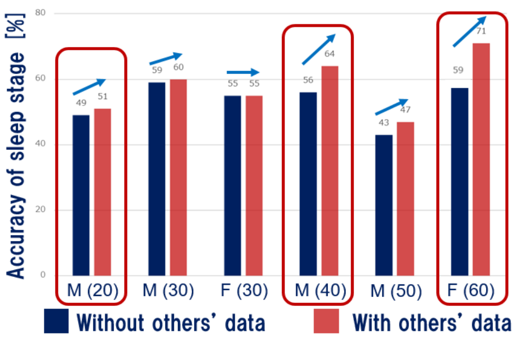


**Figure 3**: Accuracy of sleep stage estimation with/without others' data

are the same person in Figure 2 who have bad conditions, others' data can increase the accuracy of the sleep stage estimation even though own data cannot increase it.

## 4.3. Discussion

As described in Section 1, adaptability is required to cope with changing health condition. From this viewpoint, the own data and others' data contribute to improving the accuracy of the sleep stage estimation. However, own data is not useful in health condition change, while others' data which have the highest similarity of the target person is useful. This suggests that others' data has a more potential of increasing the accuracy of the sleep stage estimation than own data in health condition change.

# 5. Conclusion

This paper focused on Socially Responsible AI (SRAI) and discussed how SRAI should handle wellbeing. From the viewpoint of the AI ethics in SRAI, this paper claimed that the seven issues of the AI ethics (*i.e.*, accountability, transparency, fairness, reliability/safety, security, inclusiveness, cooperativity) are not enough for the wellbeing systems such as a healthcare system. This suggests that the "adaptability" is needed as the new concept to cope with wellbeing because health condition always changes.

To investigate the above issue, this paper focused on sleep as one of targets of the healthcare systems and conducted the human subject experiment on the sleep stage estimation. Through the experiment, the following implications have been revealed: (1) own data (*i.e.*, heartrate in this experiment) contributes to improving the accuracy of the sleep stage estimation in many cases but it is not useful in health condition change; (2) others' data with the highest similarity of the target person is useful even in health condition change, which suggests that others' data contributes to providing the "adaptability" in the sleep monitoring systems.

The following research must be done in the near future: (1) an analysis of the adaptability in other wellbeing systems; and (2) an exploration of mechanisms for adapting health condition change.

## 6. References


[1] Adadi, A., Berrada, M.: "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)", IEEE Access Vol. 6, pp. 52138–52160, 2018.

[2] Cheng, L. Varshney, K. R., Liu, H.: "Socially Responsible AI Algorithms: Issues, Purposes, and Challenges," Journal of Artificial Intelligence Research, Vol. 71, pp 1137-1181, 2021.

[3] Google: "Objectives for AI applications," https://ai.google/responsibility/principles/

[4] Harada, T., Uwano, F., Komine, T., Tajima, Y., Kawashima, T., Morishima, M., and Takadama, K: "Real-time Sleep Stage Estimation from Biological Data with Trigonometric Function Regression Model," The AAAI 2016 Spring Symposia, AAAI, pp. 348-353, 2016.

[5] Microsoft: "Responsible AI in action," https://www.microsoft.com/en-us/ai/our-approach

[6] Rechtschaffen, A. and Kales, A: "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects", *BIS/BRI University of California*, 1968.

[7] Tajima, Y., Uwano, F., Murata, A., Harada, T., and Takadama, K.: "Sleep Stage Estimation Comparing own past heartrate or other's heartrate," SICE Journal of Control, Measurement, and System Integration (JCMSI), Vol. 11, No. 1, pp. 32-39, 2018.