

1 **TITLE**

2 **Digital Twin Brain Simulator: Harnessing Primate ECoG Data for Real-Time**
3 **Consciousness Monitoring and Virtual Intervention.**

4

5 Yuta Takahashi^{1,2*}, Hayato Idei¹, Misako Komatsu³, Jun Tani⁴, Hiroaki Tomita², Yuichi
6 Yamashita^{1*}

7 1. Department of Information Medicine, National Center of Neurology and Psychiatry, Tokyo,
8 Japan

9 2. Department of Psychiatry, Graduate School of Medicine, Tohoku University, Sendai, Japan

10 3. Institution of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan

11 4. Cognitive Neurorobotics Research Unit, Okinawa Institute of Science and Technology,
12 Okinawa, Japan

13

14 ***Correspondence**

15 Yuta Takahashi

16 Email: ytakaha@ncnp.go.jp

17 Yuichi Yamashita

18 Email: yamay@ncnp.go.jp

19 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan

20 Phone number: +81-42-346-1718

21 Fax number: +81-42-346-1748

22

23 **Running Title**

24 Digital Twin Brain Simulator

1

2 **Key words**

3 Digital twin, Treatment simulation, Real-time simulation, Artificial neural network, Generative
4 artificial intelligence, Electrocorticogram, Personalized medicine

5

6 **Number of tables: 1**

7 **Number of figures: 4**

8 **Number of supplemental information: 5**

9

1 **ABSTRACT**

2 At the forefront of bridging computational brain modeling with personalized medicine, this study
3 introduces a novel, real-time, electrocorticogram (ECoG) simulator based on the digital twin brain
4 concept. Utilizing advanced data assimilation techniques, specifically a Variational Bayesian
5 Recurrent Neural Network model with hierarchical latent units, the simulator dynamically
6 predicts ECoG signals reflecting real-time brain latent states. By assimilating broad ECoG signals
7 from Macaque monkeys across awake and anesthetized conditions, the model successfully
8 updated its latent states in real-time, enhancing the precision of ECoG signal simulations. Behind
9 the successful data assimilation, a self-organization of latent states in the model was observed,
10 reflecting brain states and individuality. This self-organization facilitated simulation of virtual
11 drug administration and uncovered functional networks underlying changes in brain function
12 during anesthesia. These results show that the proposed model is not only capable of simulating
13 brain signals in real-time with high accuracy, but is also useful for revealing underlying
14 information processing dynamics.

15

1 INTRODUCTION

2 In recent years, the concept of digital twins for human organs has emerged as groundbreaking
3 technology that promises to revolutionize traditional medical frameworks [1-3]. A digital twin
4 represents an organ's functions through mathematical models, characterized by its ability to
5 synchronize with the actual organ in real-time based on observational biological signals, thereby
6 reflecting changes in the organ sequentially in the model [2, 3]. Such models, mirroring the current
7 state of an individual's organs, are anticipated to contribute to personalized treatment planning by
8 enabling pre-emptive treatment simulations [1, 2]. Moreover, the capacity to precisely simulate
9 effects of certain treatments holds promise for discovery of novel and more effective treatment
10 approaches. Applications of digital twins for human organs have been progressively realized in
11 fields such as cardiovascular disease treatment and pre-surgical simulations [1, 2]. In the fields of
12 neurological and psychiatric disorders, significant heterogeneity in single disease populations
13 underscores the heightened importance of personalized medicine, where benefits derived from
14 digital twin technologies are expected to be substantial [4]. Until recently, the complexity of the
15 brain was thought to pose significant challenges to construction of digital twins; however,
16 advances in mathematical models for brain computations, the computational infrastructure for
17 their implementation, and brain activity monitoring techniques have recently initiated embryonic
18 efforts in development of digital twin brains.

19

20 Technological contributions to digital twin brain development have included creation of brain
21 function simulators at various scales. At the microscopic scale, the focus has been on modeling
22 components of neurons, including ion channels and receptors, as well as action potentials of
23 individual neurons [5, 6]. These models offer the advantage of having a clear correspondence
24 between neuronal properties, including receptors, and model variables, allowing for simulation

1 of pharmacological effects. On the downside, modeling the whole brain based on microscopic
2 neuronal models incurs immense computational costs [6]. There is also the difficulty of non-
3 invasively acquiring microscopic data from patient brains, making it hard to incorporate real-
4 time patient information. Furthermore, models at the microscopic level are still not
5 methodologically mature enough to handle phenomena associated with higher-level cognitive
6 processes. In contrast, a considerable volume of research has focused on modeling brain
7 functions at a more macroscopic scale [7]. For instance, reinforcement learning models and
8 Bayesian inference models represent the interaction between the agent and the environment with
9 simple mathematical expressions, enabling modeling of cognitive processes as functions that
10 output behavioral choices based on sensory input from the environment [7, 8]. However, the
11 relationship between the limited number of variables that constitute these abstract models and
12 the high-dimensional biological changes induced by therapeutic interventions, such as extensive
13 changes in neuronal receptors and alterations in neural activity in various brain regions, remains
14 unclear. Consequently, using these scale models for simulating therapeutic interventions still
15 presents challenges.

16

17 In the context of bridging the gap between disparate scales of brain functionality, there has been
18 a growing interest in recent years in simulations of brain functions at the mesoscopic level [9-
19 12]. Unlike simulations that focus solely on individual neurons or local neural circuits, this
20 mesoscopic level simulator entails modeling neural activities and their interactive dynamics on
21 a regional basis in the brain. The mesoscopic level is where we can measure brain signals using
22 non-invasive techniques like electroencephalography (EEG), magnetoencephalography (MEG),
23 and functional magnetic resonance imaging (fMRI) to understand human brain function. It is
24 also where we can apply neuromodulatory methods, such as transcranial magnetic stimulation

1 and deep brain stimulation, showing how closely connected this level of modeling is to
2 advanced medical technologies. Furthermore, simulations at the whole-brain level are feasible at
3 manageable computational cost [13], making mesoscopic models promising candidates for
4 digital twin brains. However, research on digital twin brains using mesoscopic models is still in
5 nascent stages, and has yet to fully realize requirements such as real-time synchronization and
6 interventional simulations. In this endeavor, our aim is to endow digital twin brains with
7 requisite capabilities through utilization of generative artificial neural network models at the
8 mesoscopic level. Our current model's distinctive feature lies in its utilization of data-driven
9 end-to-end training, an approach that autonomously extracts information from observed brain
10 signals to accurately simulate brain functions responsible for generating these signals. To
11 further enhance the model's capacity for real-time, accurate simulations, "data assimilation" is
12 employed. This advanced technique supports data-driven optimization even after training is
13 complete, thereby enabling the model to dynamically adapt to unknown and changing
14 situations. However, the inherent complexity and noise in brain signals present a substantial
15 challenge. To address this, we employ stochastic generative models designed to capture intricate
16 probabilistic structures underlying generation of these signals. This approach allows virtual
17 therapeutic intervention to accurately simulate the process of generating changes in brain states
18 and resulting brain signals.

19

20 In this study, we examine the utility of our proposed artificial neural network model as a digital
21 twin brain, using electrocorticogram (ECoG) data. ECoG, one of intracranial EEGs, involves
22 direct placement of electrodes on the brain's surface, thereby offering higher precision in
23 measuring brain activity than traditional EEG. By conducting data-driven training from ECoG
24 signals, we aim to construct an ECoG generator (a digital twin brain prototype) and to evaluate

1 its ability to meet the requirements of a digital twin, including real-time synchronization and
2 intervention simulation, through the following procedure. Initially, we verify whether the
3 simulator can generate ECoG signals with sufficient accuracy. Considering the clinical
4 application of simulating novel patients after model development, the accuracy of generating
5 ECoG for test individuals (not used for model training) is also verified. Furthermore, we
6 evaluate the effectiveness of data assimilation, a technique to synchronize the digital twin with
7 the brain in real time. This involves validating the model's ability to progressively mirror the
8 brain's latent states, such as consciousness levels (awake and anesthetized states). Expanding on
9 these findings, we conduct virtual intervention experiments on test individuals and assess the
10 validity of the generated virtual ECoG. Additionally, by analyzing the functionality of this
11 model, we endeavor to provide a unified explanation for the underlying generation system of
12 anesthetic ECoG, a topic previously characterized by fragmented knowledge of frequencies and
13 amplitudes and an incomplete understanding of its generation mechanisms. Specifically, by
14 examining transmission of information to various brain regions in the model, we seek to define
15 the functional network required for generation of anesthetic ECoG. Furthermore, through virtual
16 intervention simulations on this functional network, we assess the contribution of each network
17 to generation of anesthetic ECoG. Finally, we investigate whether the model can express
18 individuality as a result of data-driven training and identify which variables represent this
19 individuality.

1 **RESULTS**

2 **System Overview**

3 Development of the proposed digital twin brain simulator leveraged wide-area ECoG recordings
4 of macaque monkeys, publicly accessible through the Neurotycho database [14, 15]. These data
5 consist of ECoG signals recorded from four macaques, both under anesthesia (ketamine and
6 medetomidine) and in awake, resting condition with eyes closed. Of the 128 channels covering
7 an entire hemisphere of the cerebral cortex, ECoG signals from 20 channels spanning 10 brain
8 regions were selected for use (Methods section for details). ECoG data of all four individuals
9 were utilized, with three individuals used for training and the remaining individual used for testing
10 (data assimilation), employing a four-fold cross-validation methodology.

11
12 The core component of the simulator is modeled using a Variational Bayesian Recurrent Neural
13 Network (V-RNN), which is characterized by a temporal and spatial hierarchical structure (Figure
14 1A) [16-19]. This structure comprises three levels: the local region level (the first level) operating
15 at short time scales at the lowest layer, the functional network level (the second level) functioning
16 at intermediate time scales, and the global state level (the third level) that operates over long time
17 scales at the highest layer. These level-specific dynamics are realized by latent units z , which
18 represent the stochastic latent state of each level, and deterministic units d , which convey time-
19 scale-specific deterministic dynamics.

20
21 The V-RNN operates according to a dynamic generative model acquired through data-driven
22 training (Figure 1B), executing the following procedure: Initially, it estimates hierarchical and
23 probabilistic latent states at time step t (z_t), as the “prior.” Based on this prior of z_t , the model
24 predicts the ECoG signal \hat{x}_t . Subsequently, the actual observed ECoG signal x_t is inputted. The

1 prediction error is calculated in a time window that encompasses the past spanning H time steps
2 from t ($t - H + 1 : t$). The latent state $z_{t-H+1:t}$ is updated based on this prediction error and is
3 referred to as the posterior. Predictions are then made again based upon this posterior. The
4 operation of making predictions and updating the posterior is repeated a predetermined number
5 of times. Finally, the posterior is used to generate the prior at the next time step (z_{t+1}) through
6 the deterministic state d_t . This sequential repetition of updating the latent state based on
7 predictive and observational data, a process designated as data assimilation, enables real-time
8 estimation of latent states and generation of virtual ECoG.

9
10

11 **Real-time latent state estimation via data assimilation**

12 As a result of this training, the proposed model successfully utilized acquired internal dynamics
13 to generate 20-channel ECoG sequences during both anesthetized and awake resting states.
14 Supplementary Figure S1, which plots the loss throughout the training process, demonstrates a
15 reduction in loss not only for training data, but also for validation data of test individuals. This
16 indicates that the proposed model has acquired the capability to generalize, enabling it to generate
17 ECoG sequences for test individuals.

18

19 Subsequently, the capacity of the proposed model to estimate underlying latent state changes of
20 observed ECoG signals in real-time was evaluated using ECoG data that transitions from an
21 awake state to an anesthetized state (or vice versa) in test individuals. Figure 2A presents an
22 example of test results utilizing ECoG data during the transition from wakefulness to anesthesia.
23 This figure plots observed and predicted ECoG signals (for 5 of 20 channels), their prediction
24 errors, and the time series of latent states across the first to third levels ($z^{(1)}$ to $z^{(3)}$). A comparison
25 between the observed and predicted ECoG waveforms reveals successful simulation of ECoGs

1 by the model.

2

3 Temporal changes in the latent state, reflecting data assimilation, are rapid at lower levels and
4 slow at higher levels, depending on the time scale. Notably, at the highest level ($z^{(3)}$), a switch in
5 values can be seen in response to the transition from awake to anesthetized states (as enclosed in
6 the box in Figure 2A), suggesting that the value of $z^{(3)}$ indicates global changes in brain state
7 between wakefulness and anesthesia. Dynamic alterations in these values throughout these
8 processes are depicted in a Supplementary Video.

9

10 During data assimilation, changes in latent states $z^{(3)}$ (two-dimensional) were plotted in gray on
11 the scatter plot in Figure 2B. The red and cyan points represent the values of $z^{(3)}$ for the anesthesia
12 and awake conditions, respectively, as obtained through training. These observations revealed that
13 clusters of latent states corresponding to anesthesia and awake conditions appeared to emerge.
14 The trajectory of $z^{(3)}$ across these clusters during data assimilation indicates that the model
15 accurately estimated the transition from the brain states of anesthesia to wakefulness. To
16 quantitatively validate this, data assimilation simulations were conducted using multiple ECoG
17 sequences from test individuals, including transitions from wakefulness to anesthesia and vice
18 versa. Using the silhouette width metric (Methods section for details), we evaluated whether the
19 $z^{(3)}$ states during data assimilation were appropriately positioned in clusters of wakefulness or
20 anesthesia formed during training. These results showed that the proportion of correct cluster
21 localization was, on average, 0.86 (SD 0.07) before the condition switch, and 0.73 (SD 0.17) after
22 the switch. These findings suggest that the $z^{(3)}$ states during generation of ECoG data for test
23 individuals transitioning between wakefulness and anesthesia states can estimate these conditions.

24

1 **Virtual intervention to $z^{(3)}$ (virtual drug administration)**

2 The preceding analysis revealed that estimating latent states $z^{(3)}$ from an test individual's ECoG
3 signals allows for prediction of anesthetic and awake conditions. Subsequently, we investigated
4 whether virtual ECoG waveforms, generated by setting this estimated $z^{(3)}$, accurately reflect
5 characteristics of anesthetic and awake ECoGs. This approach, with $z^{(3)}$ as a virtual intervention
6 target, is intended to impact all hierarchical levels and brain regions downstream of $z^{(3)}$, simulating
7 administration of a drug that influences the whole brain. Evaluation of virtual ECoG leveraged a
8 discriminator (3D convolutional neural network [20]) that was designed to differentiate between
9 anesthetic and awake conditions (See Methods for details.). Results demonstrated that virtual
10 ECoGs, generated by setting the estimated $z^{(3)}$ for anesthetic and awake conditions in test
11 individuals, were correctly classified into these respective conditions with high accuracies of 0.92
12 and 0.98. These results indicate that virtual ECoG signals generated by targeting $z^{(3)}$ in virtual
13 interventions accurately reflect the ECoG characteristics of both anesthetic and awake conditions.

14

15 **Functional Analysis of $z^{(2)}$ (Functional network analysis)**

16 We next explored representations that emerge in latent states of intermediate levels ($z^{(2)}$) following
17 training. The current V-RNN model incorporates multiple latent variables at the second levels
18 ($z^{(2)}$), each independently influencing deterministic units in the first level ($d^{(1)}$) corresponding to
19 various brain regions. Hence, each latent variable $z^{(2)}$ is supposed to independently govern a
20 distinct functional network. To analyze how these multiple functional networks change between
21 anesthesia and wakefulness, we calculated transfer entropy as a measure to quantify the strength
22 of control from $z^{(2)}$ s to $d^{(1)}$ s. Transfer entropy is a measure of the amount of information transfer
23 between two sequences based on information theory.

24

25 In the current V-RNN, $z^{(2)}$ s incorporates three units, whereas $d^{(1)}$ s consists of modules that

1 correspond to ten brain regions. Consequently, transfer entropy values from $z^{(2)}$ to $d^{(1)}$ in a single
2 experiment are obtained as three sets of ten-dimensional vectors. Considering the implementation
3 of cross-validation among four individuals, resulting in four separately trained networks, and
4 given two conditions, anesthesia and wakefulness, it follows that 24 vectors (3 units \times 4 networks
5 \times 2 conditions) were acquired. These vectors were subsequently sorted through hierarchical
6 cluster analysis based on Euclidean distance (Figure 3A). Figure 3A reveals that transfer entropy
7 values from $z^{(2)}$ to $d^{(1)}$ split into different clusters for anesthetized (top half of the heatmap) and
8 awake (bottom half) conditions, with a tendency for higher transfer entropy under anesthesia. This
9 indicates a stronger control mechanism from $z^{(2)}$ to $d^{(1)}$ under anesthesia than during wakefulness,
10 reflecting the emergence of more explicit functional networks in the anesthetized state. Moreover,
11 these transfer entropy values from $z^{(2)}$ to $d^{(1)}$ under anesthesia were composed of three clusters, as
12 depicted in the dendrogram colored green, yellow, and red in Figure 3A. Each cluster has been
13 designated as Functional Network (FN) 1, FN2, and FN3, respectively. These networks plausibly
14 represent the necessary networks to generate ECoG during anesthesia. Figure 3B visualizes the
15 average transfer entropy from $z^{(2)}$ to $d^{(1)}$ in FN1, FN2, and FN3 during anesthesia on the brain
16 diagram, with increased transfer entropy observed in the frontal and anterior temporal areas for
17 FN1, the frontal area and S1 for FN2, and the V1 for FN3. Figure 3C presents the transfer entropy
18 during wakefulness for each FN, showing a general decrease in transfer entropy for FN1 and FN2
19 compared to the anesthetized condition, while V1 in FN3 remains high. These results demonstrate
20 that the dynamics controlling each brain region vary between anesthesia and wakefulness, and
21 these variations are distinct for each FN.

22

23 Subsequently, the impact of these FNs on altering ECoG patterns from wakefulness to anesthesia
24 was investigated through virtual interventions targeting $z^{(2)}$. For this virtual intervention, $z^{(3)}$ states

1 were set to those of an awake condition, under which ECoG signals characteristic of wakefulness
2 would be generated in the absence of intervention. Thereafter, a virtual intervention was
3 conducted on $z^{(2)}$ states, altering each unit individually to a state representative of anesthesia. By
4 applying the ECoG signals generated through this intervention to the discriminator described in
5 the previous section, we can assess the extent to which generated ECoG signals transitioned
6 toward those characteristic of anesthesia. This approach enables evaluation of causal effects of
7 individual FNs on generation of anesthetic ECoG signals. These results intriguingly revealed
8 significant variations in the proportion of anesthetized ECoG signals generated depending on the
9 intervened FNs, with proportions of 0.12 (± 0.06) for FN1, 0.74 (± 0.25) for FN2, and 0.05 (± 0.08)
10 for FN3. These findings highlight the substantial differences in the causal effects of each FN to
11 alter ECoG signals toward an anesthetic condition.

12

13 **Functional Analysis of $z^{(1)}$**

14 Similar to the scheme outlined in the previous section, we next investigated the impact of latent
15 states at a local region level ($z^{(1)}$) on altering generation of ECoG signals from awake to
16 anesthetized conditions through a virtual intervention experiment. By setting $z^{(3)}$ s for the awake
17 condition and $z^{(1)}$ for the anesthetized condition in a targeted brain region, we generated ECoG
18 signals and used the discriminator for condition differentiation. Table 1 presents the proportion of
19 generated ECoG signals with those characteristic of anesthesia following intervention in specific
20 brain regions. The average proportion of anesthetic ECoGs ranged from 19% to 29%, with a
21 variance $\leq 10\%$ and significant individual differences (SD). While the effect size was relatively
22 small, there was a tendency for the proportion of anesthetic ECoGs to vary depending on the
23 intervention brain region. Specifically, a two-way ANOVA model, with the proportion of
24 anesthetic ECoGs as the dependent variable, and brain region and individual as independent

1 variables, indicated a trend toward a significant main effect of the brain region on the proportion
2 of anesthetic ECoGs, with a P-value of 0.064.

3

4 **Individuality analysis**

5 Finally, we assessed whether the proposed model could reflect individual characteristics, focusing
6 on $z^{(3)}$ representation. Figure 4A presents an example of the estimated probability density
7 distribution of $z^{(3)}$ states for awake and anesthetized conditions of the three individuals used in
8 training. Figures 4BCD identify which individual corresponds to each point among these training
9 results, while Figure 4E estimates the $z^{(3)}$ states of an test individual. The distribution of $z^{(3)}$ under
10 anesthesia appears to vary among individuals, whereas the distribution under awake conditions
11 appears to overlap. Quantitative evaluation based on cross-validation confirms individual-specific
12 clustering under anesthesia (average silhouette width 0.228 > 0, $P = 9.16 \times 10^{-11}$), whereas no
13 statistically significant cluster structure was observed under wakefulness (-0.09 , $P > 0.05$). Thus,
14 in terms of $z^{(3)}$ representation, individual differences are self-organized under anesthesia, while
15 individual representations are not distinctly observed under wakefulness.

16

17 Subsequently, we examined whether waveforms of virtual ECoGs generated based on the
18 estimated $z^{(3)}$ reflects individual characteristics. Evaluation of virtual ECoG leveraged a
19 discriminator (3D convolutional neural network [20]) that was designed to distinguish among the
20 four individuals used in the experiment (details in Methods). These results indicate that the
21 proportion of correctly identified ECoG signals generated for test individuals was 0.35 under
22 anesthesia and 0.30 when awake, both slightly exceeding chance levels (0.25 of selecting the
23 correct individual out of four).

1 DISCUSSION

2

3 Summary of Results

4 In this study, we developed a digital twin brain prototype capable of generating ECoG signals
5 from test individuals with high precision in real-time. This was achieved by employing a V-RNN
6 with latent variables possessing temporal and spatial hierarchical structures, which updates these
7 latent variables sequentially through data assimilation from observed ECoGs. Examination of the
8 latent variable representation that facilitated such precise simulations revealed its self-organized
9 structure at different hierarchical levels. Specifically, at higher levels of latent states ($z^{(3)}$), clusters
10 reflective of global brain states such as anesthesia and wakefulness, as well as individual
11 uniqueness, were observed. Clusters of these $z^{(3)}$ latent states enabled appropriate updates of latent
12 states during data assimilation of ECoG signals from test individuals, leading to successful virtual
13 drug administration simulations. Transfer entropy analysis between intermediate ($z^{(2)}$) and lower
14 ($d^{(1)}$) hierarchical levels indicated emergence of activity patterns that control local brain regions
15 corresponding to functional networks. Virtual interventions on these functional networks
16 demonstrated that the effect of altering generated ECoGs from wakefulness to anesthesia varied
17 significantly across functional networks.

18

19 Comparison with conventional simulators

20 There have been efforts to achieve real-time simulation of ECoG and EEG signals. For example,
21 in pursuit of EEG signal simulation, dynamics of average postsynaptic potentials (PSPs) in
22 excitatory and inhibitory interneuron populations under electrodes have been modeled using
23 differential equations [10, 21, 22]. These equations incorporate synaptic gain and time constants
24 as time-varying parameters, with these parameters being inferred in real-time from observed EEG
25 signals. Ultimately, the sum of the PSPs of excitatory and inhibitory interneurons is output as the

1 prediction of the EEG signal. While this modeling approach is notably advantageous for its low
2 computational cost, it is primarily designed for simulating a single-channel EEG signals [21],
3 posing limitations on capturing intricate interactions among various brain regions. Concomitant
4 with such real-time-focused model development, there has been an increase in attempts to create
5 models that generate interactions among brain regions, reflecting individuality based on brain
6 imaging [11, 12, 23, 24]. These efforts often employ graph network models, representing
7 interactions among brain regions with differential equations that primarily involve linear
8 transformations of state variables of brain regions. Parameters for these linear transformations,
9 such as weights or distances of connections between brain regions, are determined based on
10 diffusion brain MRI and related methods, which measures brain structures, allowing model
11 individualization [23]. These models are frequently used to simulate the BOLD signal in fMRI.
12 A notable feature of these models is that the model structure can be determined based on the actual
13 brain structure of the individual, allowing discussion of the relationship between brain structure
14 and brain function [12, 23]. On the other hand, technical challenges remain in realizing real-time
15 data assimilation. A shared attribute among these models that are currently extensively utilized is
16 a priori explicit formulation of dynamics specific to each brain signal type using differential
17 equations [21, 23]. In contrast, the primary characteristic of our model using artificial neural
18 networks is their data-driven training, offering substantial flexibility in handling input signals and
19 facilitating modeling of diverse modalities, including simultaneous modeling of multiple signal
20 types. In scenarios involving processing of brain signals, such as EEG, MEG, and fMRI, coupled
21 with sensory input signals (visual and auditory), the proposed model's general applicability could
22 be valuable. On the other hand, reduction of computational cost in the real-time PSP simulator
23 and incorporation of individual brain structures into the architecture of the graph network model
24 are notable features of these previous models, suggesting that combining insights from these

1 models can lead to refinement of the proposed model.

2

3 **Functional analysis using Digital Twin Brain**

4 The proposed model not only achieves real-time, high-accuracy simulation of brain signals, but
5 also has the potential to yield new insights into mechanisms of brain function changes during
6 anesthesia, through functional analysis. Previous studies on ketamine, a primary component of
7 the anesthesia used, have reported increased connectivity from subcortical areas, such as the
8 thalamus, to the cortex [25], and decreased connectivity between the frontal and parietal regions
9 [26, 27], particularly between M1 and S1 [28]. On the other hand, in the current study, we
10 observed increased connectivity (transfer entropy) from $z^{(2)}$ to local brain regions ($d^{(1)}$) under
11 anesthesia (Figure 3A) and found that FN2, which strongly controls brain regions M1 and S1,
12 significantly impacts generation of anesthesia ECoGs. These findings suggest an association
13 between $z^{(2)}$ and subcortical functions. Moreover, decreased connectivity between the frontal and
14 parietal regions during ketamine administration may be underpinned by enhanced control from
15 hierarchical levels higher than these brain regions, exerted by the functional network, such as FN2.
16 By this means, methodology employed in the current study could provide a new perspective,
17 supplementing and explaining background processes behind the fragmented findings of prior
18 research. Meanwhile, some findings of the current study seem to differ from previous research.
19 For instance, while prior ketamine studies have identified reductions in connectivity from the
20 frontal cortex not only to S1 but also to posterior parietal regions [27], the current study did not
21 find much change in connectivity (transfer entropy) from $z^{(2)}$ to the posterior parietal region (IPS)
22 between anesthetized and awake conditions (Figure 3A). These seemingly divergent results may
23 be attributed to differences in research methodologies. Whereas previous research focused on
24 direct inter-regional information transmission, the current study assumed hierarchical latent states

1 of top-down control over brain regions behind ECoG generation. Furthermore, unlike most prior
2 studies, the current study implemented virtual interventions in functional networks to investigate
3 their causal effects. We hope that approaches of previous studies and the current study
4 complement each other, thereby contributing to a more comprehensive understanding of complex
5 brain functions.

6

7 **Limitation**

8 In this study, we conducted a virtual anesthetic ECoG generation experiment on test individuals
9 to examine whether individual characteristics were replicated. As a result, it was confirmed that
10 individual-specific clusters emerged in higher-order latent states $z^{(3)}$; however, applying
11 discriminators to generated ECoG signals yielded an identification rate of 0.34 for these
12 individuals, which, although above chance level, remained relatively modest. Several factors may
13 account for this outcome. First, generating ECoG signals for test individuals using only training
14 data could be inherently difficult, as training data lack enough detail to accurately capture an test
15 individuals' unique characteristics. Second, optimization of network structural parameters (such
16 as the number of neurons, hierarchical levels, and connectivity) might not have been adequate,
17 suggesting that improvements in these parameters could enhance the capacity to replicate
18 individuality. In addition to improve network structures, future studies are expected to explore
19 how individuality is represented when increasing the number of training individuals or when
20 dealing with human brain signals, instead of those of macaques. Furthermore, future research is
21 expected to explore what the individual differences manifested in latent state $z^{(3)}$ represent. In this
22 study, clusters emerged only under the anesthesia condition and not under the awake condition. It
23 cannot be ruled out that these individual differences under anesthesia represent a simple
24 intervention intensity, such as the depth of anesthesia. Integrating biological signals other than

1 ECoG or information about the intervention intensity into the model inputs may facilitate a deeper
2 discussion on the relationship between intervention intensity and intrinsic individual differences.

3

4 **Brain computational theory and future perspectives**

5 The proposed model can be discussed in relation to modeling of cognitive processes from the
6 perspective of brain computation theory. The model executes free energy minimization through
7 parameter optimization, which is a core aspect of the free energy principle derived from Bayesian
8 brain theory [29, 30]. This theory hypothesizes that the brain's information processing is
9 analogous to data assimilation. Indeed, similar RNN models have been utilized in computational
10 modeling and neurodevelopmental robotics research to understand cortical information
11 processing, cognitive functions, and psychiatric symptoms [31-35]. The present research may be
12 interpreted as demonstrating that the free energy principle-based approach is beneficial for
13 simulating not only cognitive processes, but also brain signals. Future work is anticipated to
14 expand this model to simultaneously simulate cognitive processes, including sensory inputs and
15 motor outputs, in addition to brain signals, including ECoG, EEG and fMRI. This approach
16 indicates the potential to uncover latent state changes underlying cognitive features in
17 neuropsychiatric disorders and to facilitate virtual intervention experiments. Consequently, the
18 model proposed here, with its ability to encapsulate multi-level phenomena, holds promise to
19 illuminate brain information processing, potentially serving as a prototype for a digital twin brain.

1 **Methods**

2 **Data and preprocessing**

3 In this study, we utilized ECoG data from macaques, sourced from the publicly available
4 Neurotycho database[14, 15, 36, 37]. To reduce computational costs, we selected a total of 20
5 channels, with two channels from each of the ten representative brain regions, from the entire
6 set of 128 ECoG channels distributed across the hemisphere. These representative regions
7 include the frontal pole (FP), primary visual cortex (V1), higher visual cortex (HV), auditory
8 cortex (AC), primary somatosensory cortex (S1), primary motor cortex (M1), intraparietal
9 sulcus (IPS), premotor cortex (PM), dorsolateral prefrontal cortex (DLPFC), and anterior
10 temporal cortex (AT). ECoG signals were recorded at 1000 Hz. We conducted experiments
11 using a cross-validation approach (training on three individuals, testing on one) with four
12 individuals (George, Chibi, Kin2, Su), employing data collected during both under anesthesia
13 (ketamine and medetomidine) and in an awake, resting condition with eyes closed.

14

15 Preprocessing of ECoG data encompassed re-referencing, outlier exclusion, and linear
16 normalization. During the re-referencing process, the Common Median Reference method was
17 applied. Outlier exclusion adhered to protocols from the preceding study using the same dataset
18 [15], where data were divided into 2000 time-step (2-second) bins, and bins with values
19 exceeding eight standard deviations, as well as their immediate neighbors, were excluded.
20 Consequently, 1.4% of data under anesthesia and 8.6% in awake macaques were excluded. For
21 linear normalization, a scale transforming the range of values across all subjects, examination
22 days, and experimental conditions to [-1, 1] was implemented.

23

1 Following these preprocessing steps, data were segmented again into 2000 time-step (2-second)
2 intervals from which twelve anesthetized and twelve awake sequences were randomly selected
3 from each of the three training individuals, totaling 72 sequences (144000 time steps, 144
4 seconds) used for training. Then, to validate our model's efficacy on test individuals' ECoG data,
5 we devised three sets of test data for different experiments. For the real-time latent state
6 estimation experiment, we crafted 25 sequences by concatenating ECoG signals: the first 2500
7 time steps were drawn from randomly selected awake sequences, whereas the latter 1500 time
8 steps were derived from randomly chosen anesthetized sequences. Another 25 sequences were
9 similarly prepared, but with the order of states reversed, totaling 50 sequences. This disparity in
10 time steps between the first and second halves accounts for consideration of a time window
11 width H of 500 time steps. For the virtual intervention experiment to $z^{(3)}$ and individuality
12 analysis, we randomly selected 100 sequences of ECoG signals truncated at 2000 time steps, 50
13 under anesthesia and 50 while awake. Similarly, for virtual intervention experiments on $z^{(2)}$ or
14 $z^{(1)}$, we randomly selected 50 sequences of ECoG signals truncated at 2000 time steps, 25 under
15 anesthesia and 25 while awake.

16

17 **System Overview**

18 This simulator's main component employed a V-RNN, characterized by its spatial and temporal
19 hierarchy. This hierarchy consists of the following three levels: a local region level (the first
20 level), corresponding to modules for each of the 10 brain regions, operating on rapid changes at
21 short time scales; the functional network level (the second level), consisting of a single module
22 that controls local region modules and operates on slower changes over longer time scales; and
23 the global state level (the third level), also a single module, which maintains constant latent
24 states during generation of sequence predictions to map to sequential patterns of the target

1 ECoG at an abstract level. These hierarchical structures are facilitated by each module
2 comprising probabilistic latent units z and deterministic units d with time-scale-specific
3 dynamics. In this model, generation of ECoG signals is characterized by top-down prediction,
4 where information flows from latent units z to deterministic units d , descending to lower
5 levels. Dynamics in these top-down predictions are acquired through a data-driven training
6 phase. Following the training phase, a data assimilation phase utilizing ECoG data from test
7 individuals facilitates real-time simulations. In the data assimilation process, initially, the latent
8 states z_t at each hierarchical level for time step t are estimated as “priors”. Leveraging these
9 priors, the model conducts top-down prediction of the ECoG signals \hat{x}_t . Subsequently, the
10 actual observed ECoG signal x_t is inputted. Based on the prediction error (the difference
11 between \hat{x}_t and x_t), latent states z_t are updated, enabling a bottom-up estimation of the
12 “posterior”. This posterior informs the next latent state z_{t+1} estimation, which becomes the
13 new prior. This iterative process of top-down prediction and bottom-up posterior updates
14 enables real-time estimation of latent states and the generation of virtual ECoG signals.

15

16 **Top-down prediction generation**

17 In top-down prediction, the prior and posterior of latent states z are utilized to derive
18 deterministic states d , enabling ECoG signal prediction \hat{x} . To illustrate this process, we
19 initially discuss calculation of the prior for latent states z . Specifically, we consider the prior
20 distribution $p(z_{t,i}^{(s)})$ for the i th latent state $z_{t,i}^{(s)}$ at time step t of the s th target sequence, at
21 local region and functional network levels. This prior distribution is assumed to follow a
22 Gaussian distribution, where the mean $\mu_{t,i}^{(s),p}$ and standard deviation $\sigma_{t,i}^{(s),p}$ are derived from
23 the previous deterministic states $d_{t-1,j}^{(s)}$.

$$p(z_{t,i}^{(s)}) = p(z_{t,i}^{(s)} | d_{t-1,j}^{(s)}) = \mathcal{N}(z_{t,i}^{(s)}; \mu_{t,i}^{(s),p}, \sigma_{t,i}^{(s),p}) \quad (1)$$

$$\mu_{t,i}^{(s),p} = \tanh \left(\sum_j w_{ij} d_{t-1,j}^{(s)} \right) \quad (2)$$

$$\sigma_{t,i}^{(s),p} = \exp \left(\sum_j w_{ij} d_{t-1,j}^{(s)} \right) \quad (3)$$

$$z_{t,i}^{(s)} = \mu_{t,i}^{(s),p} + \sigma_{t,i}^{(s),p} \times \varepsilon \quad (4)$$

1 Here, j is an index representing deterministic units in the same module as $z_{t,i}^{(s)}$. In Equation
 2 (4), ε is sampled from the $\mathcal{N}(1,0)$ distribution to obtain the latent state $z_{t,i}^{(s)}$. As described
 3 above, the latent state $z_{t,i}^{(s)}$ changes at each time step at local region and functional network
 4 levels, but at the global state level, the $z_{t,i}^{(s)}$ is kept constant during prediction generation.
 5 Therefore, at the global state level, $z_{t,i}^{(s)}$ has a prior distribution of $\mathcal{N}(1,0)$ only at the initial
 6 time step ($t=1$).

7

8 Next, the posterior distribution $q(z_{t,i}^{(s)})$ of the latent states is calculated as the conditional
 9 probability dependent on the prediction error signals $e_{t:T}^{(s)}$.

$$q(z_{t,i}^{(s)} | e_{t:T}^{(s)}) = \begin{cases} q(z_{t-1,i}^{(s)} | e_{t-1:T}^{(s)}) = q(z_{1,i}^{(s)} | e_{1:T}^{(s)}) = \mathcal{N}(z_{1,i}^{(s)}; \mu_{1,i}^{(s),q}, \sigma_{1,i}^{(s),q}) (i \in I_{GZ}), \\ \mathcal{N}(z_{t,i}^{(s)}; \mu_{t,i}^{(s),q}, \sigma_{t,i}^{(s),q}) (i \in I_{FZ}, I_{L(n)Z}) \end{cases} \quad (5)$$

$$\mu_{t,i}^{(s),q} = \tanh(a_{t,i}^{(s),\mu}) \quad (6)$$

$$\sigma_{t,i}^{(s),q} = \exp(a_{t,i}^{(s),\sigma}) \quad (7)$$

$$z_{t,i}^{(s)} = \mu_{t,i}^{(s),q} + \sigma_{t,i}^{(s),q} \times \varepsilon \quad (8)$$

1 Here, $T^{(s)}$ represents the length of the s th target ECoG sequence. $a_{t,i}^{(s)}$ is defined as the unit's
 2 adaptive internal state that represents the posterior distribution. $a_{t,i}^{(s)}$ is dynamically updated at
 3 each timestep in every target ECoG sequence throughout the processes of learning and data
 4 assimilation. This updating process is influenced by the prediction error signal $e_{t:T}$, which is
 5 propagated via a back-propagation-through-time (BPTT) algorithm. Thus, as inferred from
 6 equations (6) and (7), the posterior distribution of latent states may be conceptualized as a
 7 prediction-error-related unit state.

8

9 Based on calculated latent states $z_{t,i}^{(s)}$ as described above, internal states $h_{t,i}^{(s)}$ and outputs $d_{t,j}^{(s)}$
 10 of deterministic units are computed as follows. This process involves deriving $h_{t,i}^{(s)}$ from the
 11 previous deterministic output $d_{t-1,j}^{(s)}$ and latent states $z_{t,j}^{(s)}$ in the same module, alongside
 12 inputs from a higher-level module (Supplementary figure S2).

$$h_{t,i}^{(s)} = \begin{cases} \frac{1}{\tau_i} \left(\sum_{j \in I_{Fd}} w_{ij} d_{t-1,j}^{(s)} + \sum_{j \in I_{Fz}} w_{ij} z_{t,j}^{(s)} + \sum_{j \in I_{Gz}} w_{ij} z_{t,j}^{(s)} + b_i \right) + \left(1 - \frac{1}{\tau} \right) h_{t-1,i}^{(s)} & (i \in I_{Fd}) \\ \frac{1}{\tau_i} \left(\sum_{j \in I_{L(n)d}} w_{ij} d_{t-1,j}^{(s)} + \sum_{j \in I_{L(n)z}} w_{ij} z_{t,j}^{(s)} + \sum_{j \in I_{Fd}} w_{ij} d_{t,j}^{(s)} + b_i \right) + \left(1 - \frac{1}{\tau} \right) h_{t-1,i}^{(s)} & (i \in I_{L(n)d}) \end{cases} \quad (9)$$

$$d_{t,j}^{(s)} = \tanh(h_{t,i}^{(s)}) \quad (i \in I_{Fd}, I_{L(n)d}) \quad (10)$$

13 Here, $I_{L(n)d}, I_{Fd}, I_{Gd}$ represent index sets for deterministic units at the local region level,
 14 functional network level, and global state level, respectively, while $I_{L(n)z}, I_{Fz}, I_{Gz}$ represent
 15 index sets for latent units at their corresponding levels. The $L(n)$ takes on $n = 1, 2, \dots, 10$,
 16 corresponding to modules representing ten local brain regions. The w_{ij} refers to the weight of
 17 synaptic connections between units, and τ_i denotes the time constant. Deterministic units with

1 smaller time constants τ_i tend to exhibit rapid changes in states, whereas units with larger time
 2 constants τ_i show a tendency for slower changes in states[19]. By setting $\tau_i = 2$ at the local
 3 region level and $\tau_i = 4$ at functional network level, the model represents changes on a shorter
 4 timescale at the lower level and on a longer timescale at the higher level, effectively
 5 constructing a temporal hierarchical structure.

6
 7 Lastly, the prediction $\hat{x}_{t,i}^{(s)}$ of ECoG signals for 20 channels corresponding to 10 brain regions
 8 is generated by the respective local brain region module.

$$\hat{x}_{t,i}^{(s)} = \tanh \left(\sum_{j \in I_{L(n)o}} w_{ij} d_{t,j}^{(s)} \right) (i \in I_{L(n)o}) \quad (11)$$

9 Here, the $I_{L(n)o}$ denotes index sets of output units in the brain region $L(n)$.

11 **Parameter updates through bottom-up optimization**

12 Optimization processes in both training and data assimilation phases have been implemented by
 13 employing the negative Evidence Lower Bound (ELBO) as the loss function. This negative
 14 ELBO is equivalent to the variational free energy [38] in the context of brain computation
 15 theory, and is hereafter identified as F_t . In the framework of V-RNN, F_t is derived as
 16 follows[18].

$$F_t = \frac{1}{2} (x_t - \hat{x}_t)^2 + \sum_{l=1}^3 W^{(l)} D_{KL} [q(z_t^{(l)} | e_{t:T}) \| p(z_t^{(l)} | d_{t-1}^{(l)})] \quad (12)$$

17 The first term of this equation represents the prediction error, while the second term denotes the
 18 influence of the prior $p(z_t^{(l)} | d_{t-1}^{(l)})$ on the update of the posterior $q(z_t^{(l)} | e_{t:T})$ of the latent
 19 state. Consequently, in an optimization process aimed at minimizing F_t , the update of the
 20 posterior progresses in a manner that minimizes the prediction error under the influence of the

1 prior. The hyperparameter $W^{(l)}$, referred to as the meta-prior[17], serves to balance the
2 prediction error and the divergence between the prior and posterior, and can be individually set
3 for each hierarchical level l . In this study, we set the meta-prior to a uniform value across all
4 levels ($W^{(1)} = W^{(2)} = W^{(3)} = 0.001$). This optimization by minimizing F_t was carried out
5 during both the training and data assimilation phases.

6

7 During the training phase, synaptic weights w and the posterior are updated to minimize the
8 free energy $F_t^{(s)}$ over all time steps and target sequences.

$$F = \sum_{s \in I_s} \sum_{t=1}^{T^{(s)}} F_t^{(s)} \quad (13)$$

9 On the other hand, during the data assimilation phase, only the posterior is updated, while the
10 weights w remain fixed. In this phase, the free energy is summed over the time window
11 spanning the current time step t backward through H time steps.

$$F = \sum_{t'=t-H+1}^t F_{t'} \quad (14)$$

12 Utilizing the summed free-energy, the posterior of $z_{t-H+1:t}$ in time windows of all modules is
13 updated, with this window advancing over time. In both training and data assimilation phases,
14 the Rectified Adam optimizer was employed for parameter updates [39]. Partial derivatives of
15 the free energy with respect to each optimization parameter were computed using the BPTT
16 algorithm.

17

18 **Data assimilation**

19 Following the training phase, the data assimilation process is applied to ECoG signals of test
20 individuals. Data assimilation involves a cyclical process of bottom-up posterior updates

1 informed by top-down predictions and observations. This process occurs in a sliding time
2 window that spans from the current time step t backward through H time steps, and
3 incrementally moves forward as time progresses (Supplementary Figure S3). The process
4 unfolds as follows: (1) The prior of z_t is estimated based on the posterior of the previous time
5 step. (2) Based on this prior of z_t and the posterior at earlier time steps ($z_{z-H+1:t-1}$), ECoG
6 signals $\hat{x}_{t-H+1:t}$ in the time window are predicted in a top-down manner. (3) Subsequently,
7 new observed ECoG signals x_t are inputted. (4) At each time step in the time window, the free-
8 energy F is calculated based on the prediction error (the discrepancy between x and \hat{x}). (5)
9 To minimize the free-energy F , latent states $z_{t-H+1:t}$ are updated as the posterior. (6) Based
10 on this updated posterior, top-down predictions for $\hat{x}_{t-H+1:t}$ are performed again. This process
11 of bottom-up posterior updates (4)(5) and top-down prediction (6) is repeated a predetermined
12 number of times. Subsequently, the process returns to (1), where, from the posterior of z_t , the
13 prior for the next time step z_{t+1} is estimated. Such data assimilation enables real-time
14 estimation of latent states and generation of virtual ECoG signals.

15

16 **Parameter setting**

17 Based on preliminary investigations, the number of latent units z and deterministic units d
18 have been configured to accurately predict ECoG signals. At the local region level, for each
19 module in a respective brain region, one z and fifteen d units are allocated. At the functional
20 network level, three z and fifteen d units; at the global state level, two z units. Note that at
21 the global state level, the z states are set to remain constant during prediction generation and d
22 units are not used. During the training phase, synaptic weights w and the posterior were
23 updated 200,000 times. In the data assimilation phase, specifically during real-time latent state
24 estimation experiments, the time window width H was set to 500 time steps, with the posterior

1 updated 100 times at each timestep. For virtual intervention experiments, the time window
2 width H was fixed to the sequence length (2000 time steps), and the posterior was updated
3 10,000 times. The reason for fixing time window to sequence length here is to obtain a
4 representation of $z^{(3)}$ that abstractly corresponds to characteristics of the full ECoG sequence
5 length by not allowing $z^{(3)}$ to change temporally. Hyperparameters of Rectified Adams during
6 learning and data assimilation phases were set with $\alpha = 0.001$ (learning rate), $\beta_1 = 0.9$, and $\beta_2 =$
7 0.999.

8

9 **Virtual intervention**

10 Virtual intervention experiments were performed on latent states of $z^{(3)}$, $z^{(2)}$, $z^{(1)}$. In these
11 experiments, we generated virtual ECoGs by inputting latent states estimated as anesthetized or
12 awake test individuals in target latent units. Then, we investigated whether resulting virtual
13 ECoG waveforms exhibited characteristics of anesthetized or awake conditions. Latent states
14 used as intervention inputs were posterior distributions obtained through data assimilation using
15 ECoG signals of test individuals (See also the section of Parameter setting).

16

17 In virtual interventions targeting $z^{(3)}$, the posterior of the constant $z^{(3)}$ was input to latent units
18 together with the posterior of $z^{(2)}$ and $z^{(1)}$ to generate virtual ECoGs. Effects of this virtual
19 intervention were evaluated based on the proportion of generated ECoGs indicating anesthesia
20 or awake characteristics. Assessment of these characteristics in ECoG signals utilized a
21 discriminator, as described below.

22

23 As part of functional analysis, virtual intervention targeting $z^{(2)}$ and $z^{(1)}$ were conducted. This
24 was undertaken to investigate the contribution of each latent unit in altering generated ECoGs

1 from awake to anesthetized conditions. Initially, the posterior of an awake condition was input
2 to $z^{(3)}$ so that awake ECoGs were generated in the absence of intervention. Subsequently, the
3 posterior of anesthesia condition was individually input to specific latent units of $z^{(2)}$ or $z^{(1)}$
4 (virtual intervention) and virtual ECoGs were generated. The contribution of each latent unit to
5 anesthetized ECoG generation was evaluated based on the proportion of generated ECoG
6 manifesting anesthetic characteristics, which was assessed by the discriminator, as described
7 below. In these analyses, target latent units had their posterior distributions input, while non-
8 target latent units underwent top-down prediction generation using prior distributions without
9 explicit setting of distributions corresponding to awake or anesthetized conditions.

10

11 **Discriminator**

12 In order to examine the characteristics of virtually generated ECoG signals, a discriminator was
13 used. This discriminator made classifications as to whether the virtual ECoG signals pertained
14 to states of anesthesia or wakefulness, or in which of the four individuals ECoG signals
15 originated. The discriminator employed a model that converts ECoGs into spectrograms and
16 utilizes a deep convolutional neural network (CNN) for classification, noted for its high
17 accuracy in previous studies[40]. Training for this model utilized ECoG signals across all
18 individuals in the study, during both anesthesia and wakefulness. Generation of spectrograms
19 was carried out by applying the multi-taper time-frequency spectrum method to ECoG
20 sequences that were segmented every 2000ms. During this process, the length of the moving
21 window was set to 100ms, with a step size of 10ms, and calculations were performed for the 0-
22 200Hz frequency band. Spectrogram images across 20 channels were compiled as a single-data
23 unit and subjected to discriminant analysis using a three-dimensional CNN, the structure of
24 which is detailed in Supplementary Figure S4. In the training phase of this CNN, cross-entropy

1 was employed as the loss function, and the model was updated 100 times using an Adam
2 optimizer with a learning rate of 0.0005. Models were constructed for both binary classification
3 tasks, distinguishing between anesthesia and wakefulness, and for a four-class identification
4 task involving individual differentiation. Test accuracies for these models were remarkably high,
5 at 99.6% and 98.1%, respectively.

6

7 **Silhouette analysis**

8 The cluster structures emerging in the distribution of latent states $z(3)$ at a global state level
9 were quantitatively assessed. For this assessment, silhouette width [41], a method commonly
10 used for cluster evaluation, was used for the $z(3)_\mu$ space as shown in Figure 3A, and the
11 calculation process is shown below. Consider the case where data point i corresponding to one
12 $z(3)_\mu$ state (corresponding to one point in Figure 3A) is to be evaluated as to whether it
13 belongs to cluster C_i . To begin with, $a(i)$ is defined as follows.

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (15)$$

14 Here, $|C_i|$ is the number of points belonging to cluster C_i , and $d(i, j)$ is the distance between
15 data points i and j . $a(i)$ represents the average of the distances from data point i to the rest
16 of the points in its cluster C_i , indicating that a reduced $a(i)$ value reflects greater cohesion in
17 cluster C_i . Next, $b(i)$ is defined as follows.

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (16)$$

18 Here, $b(i)$ represents the average distance between data point i and points in the adjacent, but
19 distinct cluster, indicating that a higher $b(i)$ value suggests a greater separation from adjacent
20 clusters.

21

1 Based on $a(i)$ and $b(i)$, the silhouette width $s(i)$ for data point i is calculated as follows.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (17)$$

2 A higher value of $s(i)$ indicates that data point i is more strongly clustered in C_i . In real-time

3 latent state estimation experiments, we utilized $s(i)$ to classify each time step's $z(3)\mu$,

4 obtained through data assimilation, as belonging to either the anesthesia or awake clusters

5 established during the learning process. Specifically, the $s(i)$ was calculated for both the

6 anesthesia and awake clusters, and data i was classified into the cluster with larger $s(i)$.

7

8 From equation (17), it can be inferred that if $s(i) > 0$, data point i tends to be clustered into C_i ,

9 rather than being randomly distributed. In this study, we utilized this principle to investigate

10 whether individual clusters emerge in the $z(3)$ space. Specifically, the individual to which each

11 data point i belongs were designated as C_i and $s(i)$ was calculated for all data points in the

12 training ECoG sequences to statistically examine whether the average silhouette width exceeds

13 zero (t-test).

14

15 **Data Availability**

16 ECoG data utilized in this study is available at: <http://www.neurotycho.org/>

17

18 **Code Availability**

19 The code for the V-RNN, a principal component of the model discussed herein, is shared at:

20 https://github.com/h-idei/pvrnn_sa

21

22 **Acknowledgements**

1 This work was supported by JSPS KAKENHI (JP21K15723, JP24K20897, JP22KJ3167,
2 JP20H00625, JP24H00076, JP24K00499, JP24H02175), JST CREST (JPMJCR21P4), JST
3 Moonshot R&D (JPMJMS2031), AMED (JP21tm0424601), and Intramural Research Grant (4-
4 6, 6-9) for Neurological and Psychiatric Disorders of NCNP.

5

6 **Author contribution statement**

7 YT and YY conceived the study, HI contributed to model building, YT conducted experiments
8 and analysis, YT prepared the original draft, and YT, HI, MK, JT, HT and YY conducted peer
9 review and editing.

10

1 Table 1. Proportion of anesthetic ECoGs generated by intervention brain region

Region	FP	DLPFC	PM	M1	AT	AC	HV	S1	IPS	V1
Mean	0.21	0.23	0.24	0.19	0.22	0.26	0.27	0.28	0.29	0.21
SD	0.32	0.34	0.31	0.29	0.34	0.37	0.39	0.34	0.36	0.33

2 Abbreviations: SD, standard deviation; FP, frontal pole; DLPFC, dorsolateral prefrontal cortex;

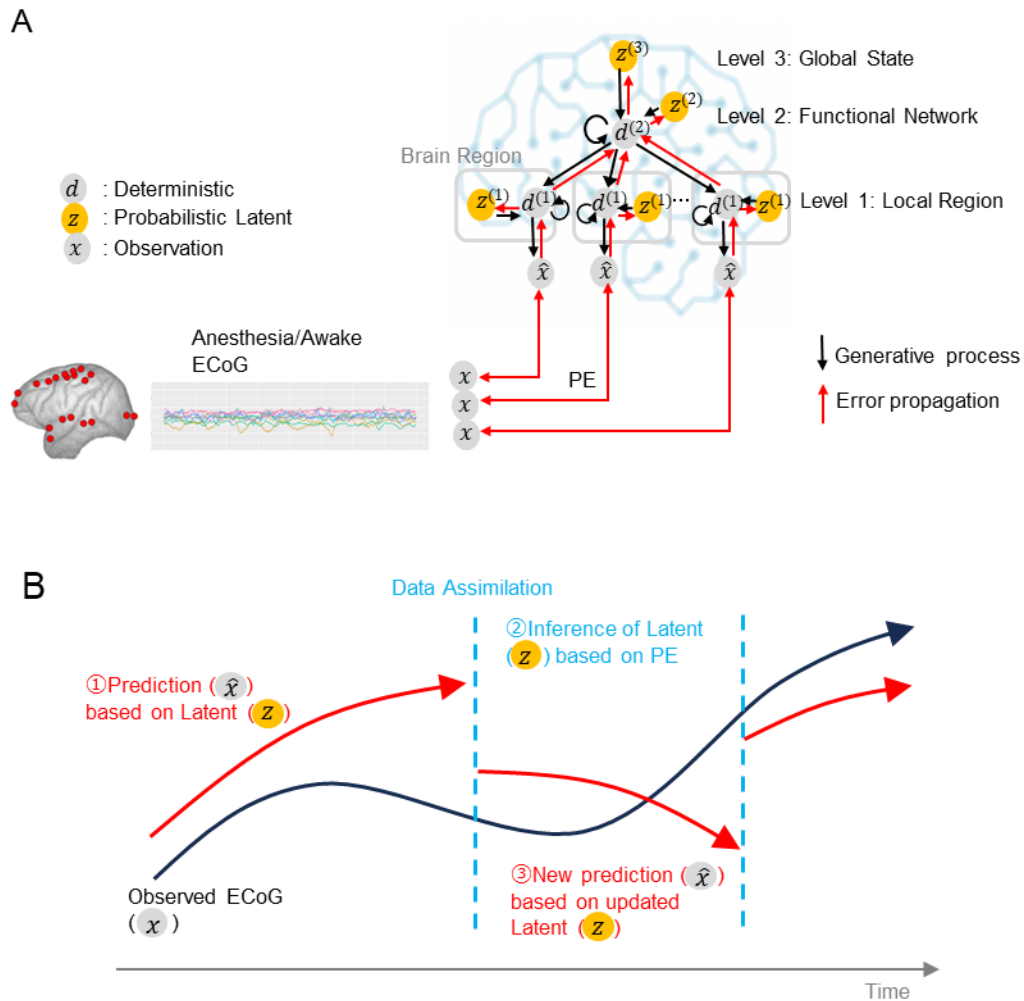
3 PM, premotor cortex; M1, primary motor cortex; S1, primary somatosensory cortex; IPS,

4 intraparietal sulcus; AT, anterior temporal cortex; AC, auditory cortex; HV, higher visual cortex;

5 V1, primary visual cortex.

6

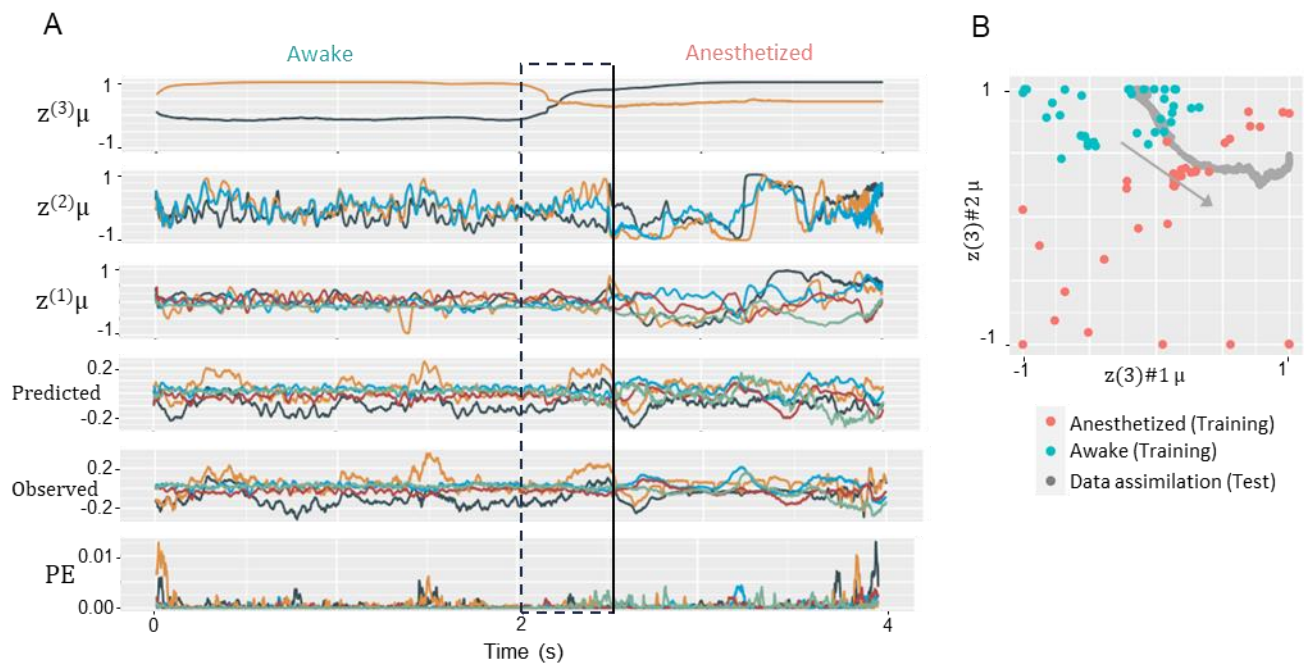
1



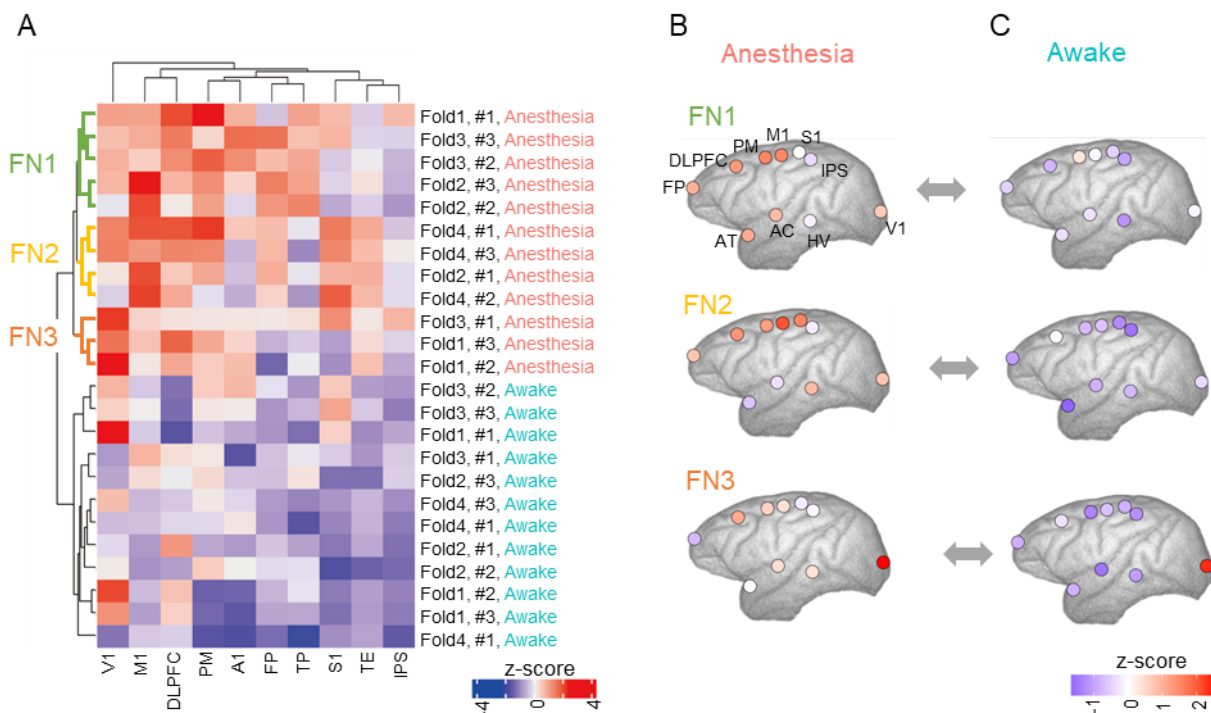
2

Figure 1. (A) Model Overview. This model depicts the process by which ECoG signals are generated from latent states z of the brain. The input x consists of wide-area ECoG measurements from 20 channels across 10 brain regions in macaques. The latent variable z possesses three hierarchical levels, each of which is stochastic. Latent states z , through synaptic connections of deterministic units d , lead to generation of ECoG signals \hat{x} . (B) Overview of the data assimilation procedure. For simplification, this figure does not show the time window. In an incrementally advancing window, predictions based on latent variables and updates of latent states based on prediction error (PE) are alternately repeated for a number of iterations.

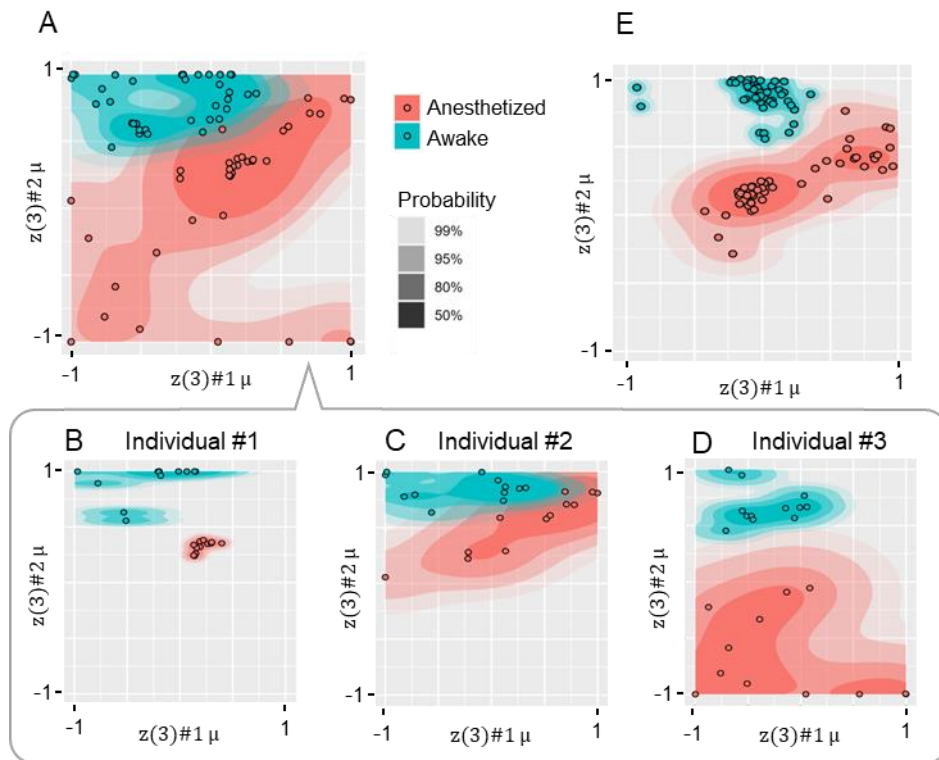
10



1 **Figure 2.** (A) Prediction and latent state sequences in data assimilation. This figure illustrates
2 the sequence of predictions and the mean (μ) of posterior distributions of latent states $z^{(1)}$, $z^{(2)}$,
3 $z^{(3)}$, generated in real-time in response to observed ECoG data. The Observed ECoG combines
4 2500 time steps (2.5 seconds) of awake ECoG observations followed by 1500 time steps (1.5
5 seconds) of anesthetized ECoG observations from a test individual (an individual not included
6 in the training dataset). This figure plots values after the data assimilation process is complete
7 across all time steps of the observed ECoG sequence. The vertical solid line shows the moment
8 of transition between awake and anesthetized conditions in observed ECoG data, with dashed
9 lines showing the time window's range when reaching this moment. In this range, updates of
10 latent states are made in response to the transition in the observed ECoG. (B) Example of the
11 distribution of the mean (μ) of $z^{(3)}$ obtained during training and data assimilation phases. In this
12 study, the latent unit $z^{(3)}$ is prepared in two dimensions, and the x- and y-axes represent the $z^{(3)}\mu$
13 values in the first and second dimensions, respectively. The gray arrow indicates the direction in
14 which $z^{(3)}$ changed over time during data assimilation.



1 **Figure 3.** Transfer entropy analysis.
2 (A) This heatmap represents transfer entropy values from $z^{(2)}$ latent variables to $d^{(1)}$ in various
3 brain regions. Rows in the heatmap correspond to $z^{(2)}$ latent variables used in calculation of
4 transfer entropy. These rows represent 24 unique $z^{(2)}$ states, derived from four variations in the
5 training dataset during cross-validation (denoted as Fold 1-4), three $z^{(2)}$ latent variables per one
6 model (denoted as #1-3), and two conditions: anesthetized and awake. Columns indicate brain
7 regions associated with $d^{(1)}$ activities utilized in transfer entropy calculations. Hierarchical
8 clustering was performed on rows and columns of this heatmap matrix. In this clustering,
9 individual elements build a hierarchy by progressively merging clusters based on Euclidean
10 distances. The heatmap color scale with transfer entropy values (z-score) from each $z^{(2)}$ to the
11 corresponding brain region $d^{(1)}$, indicating the extent of control (or information flow). (B)(C)
12 Mapping of transfer entropy values from $z^{(2)}$ to $d^{(1)}$ in different brain regions for FN1, FN2, and
13 FN3.
14 Abbreviations: FP, frontal pole; DLPFC, dorsolateral prefrontal cortex; PM, premotor cortex;
15 M1, primary motor cortex; S1, primary somatosensory cortex; IPS, intraparietal sulcus; AT,
16 anterior temporal cortex; AC, auditory cortex; HV, higher visual cortex; V1, primary visual cortex.



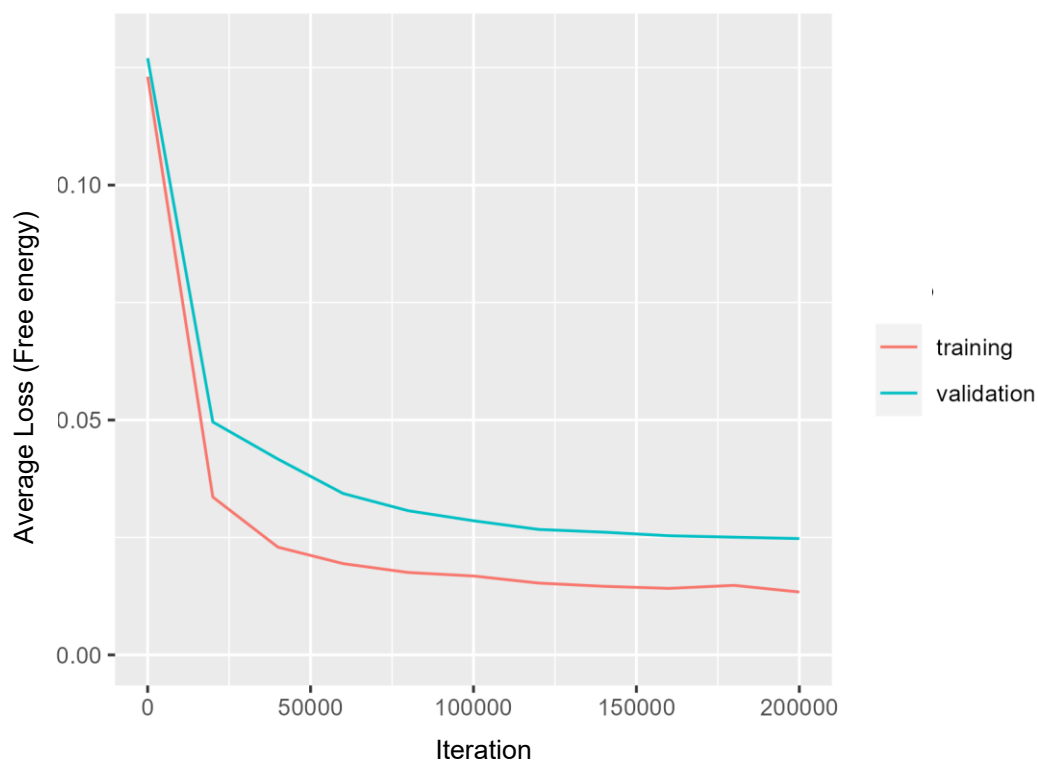
1

2 **Figure 4.** (A) Distribution of $z^{(3)}\mu$ post-training, as shown in Figure 2B, overlaid with an estimated
3 probability density plot. Each layer of varying shades represents boundaries of corresponding
4 probability levels. Estimation of bivariate probability density utilized the kernel density estimator
5 from the “ggdensity” package in R.

6 (B) (C) (D) Distributions of $z^{(3)}\mu$ depicted in Figure 3A are shown separately for each of three
7 individuals in the training data.

8 (E) Shows the distribution of $z^{(3)}\mu$ estimations from data assimilation applied to ECoG sequences
9 from a test individual.

1 Supplementary Figures

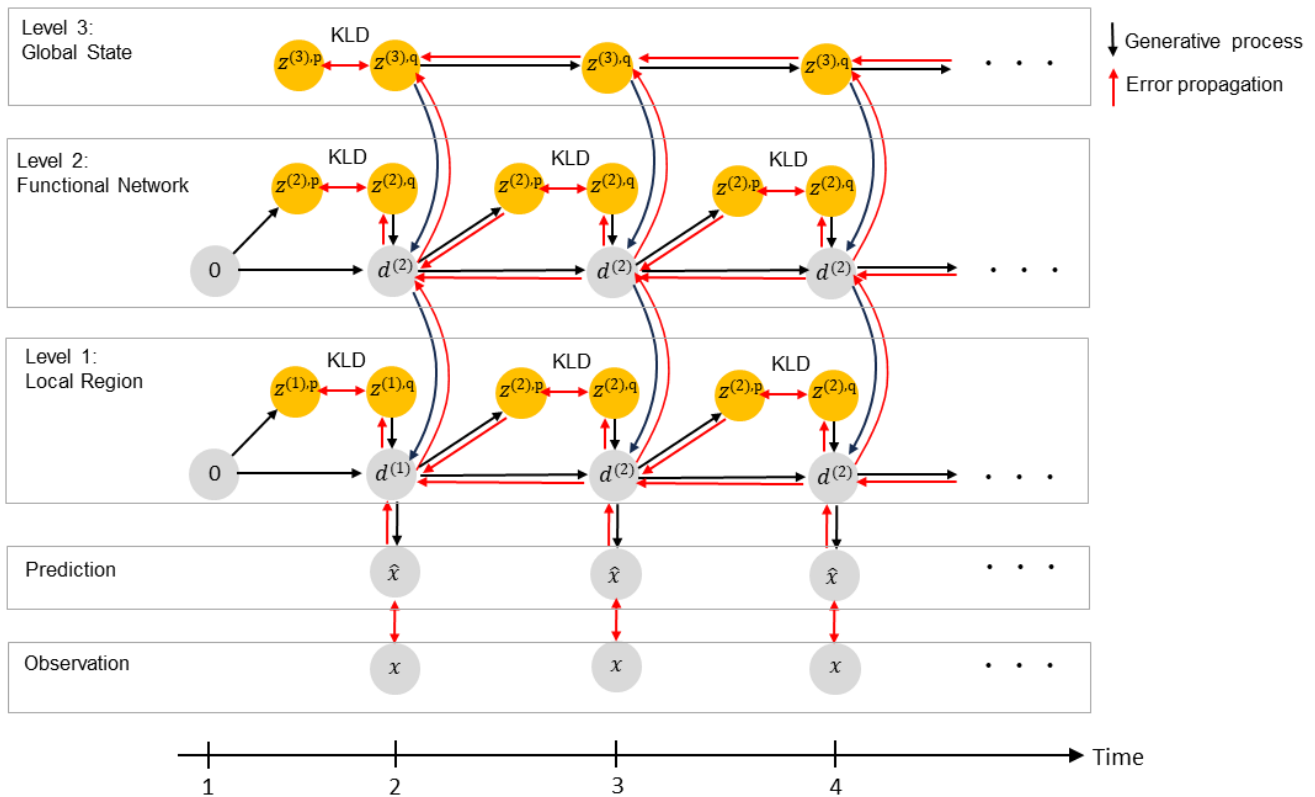


2

3 Supplementary Figure S1: Learning curve.

4

1

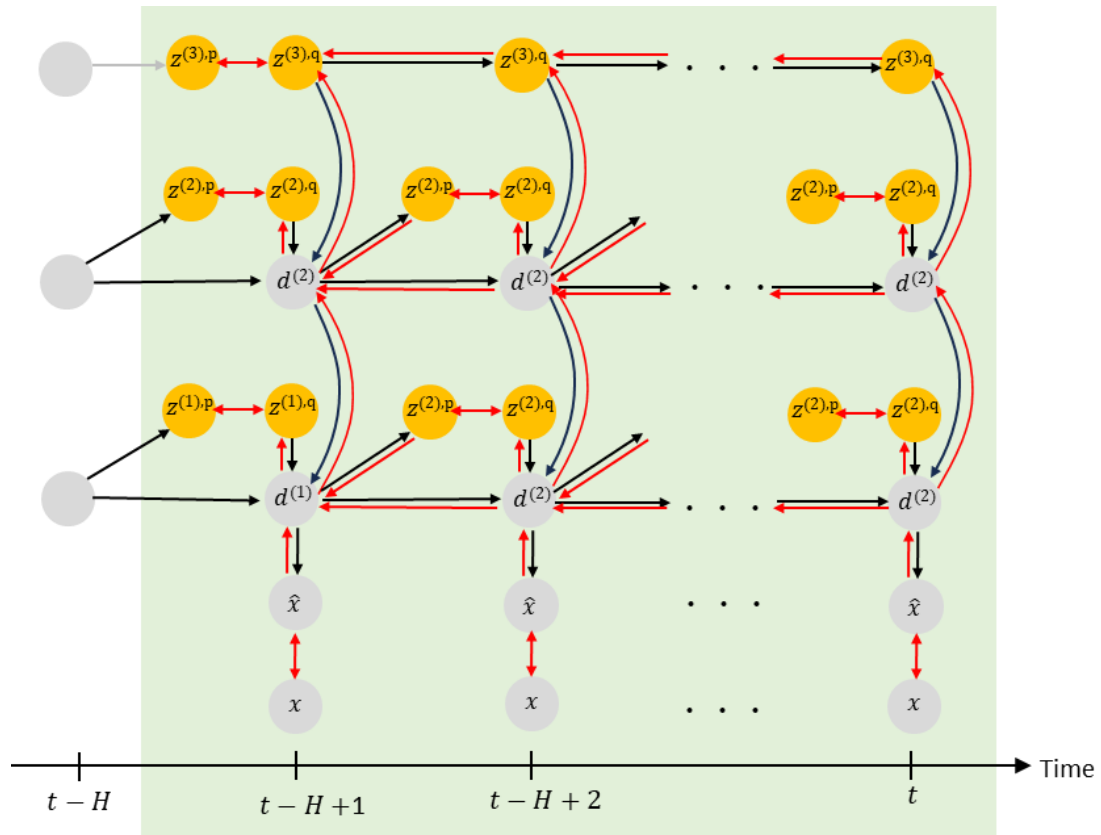


2

3 Supplementary Figure S2: Temporal processing of the V-RNN.

4 During the training phase, the network optimizes the posteriors of latent variables $z_{1:T}$ across all
 5 modules and updates time-constant synaptic weights by minimizing the free energy F over the
 6 duration T of training ECoG sequences. Initial deterministic states of all modules are set to zero.
 7 For simplicity, only one out of ten modules is depicted for local region modules. Abbreviation:
 8 KLD, Kullback-Leibler divergence.

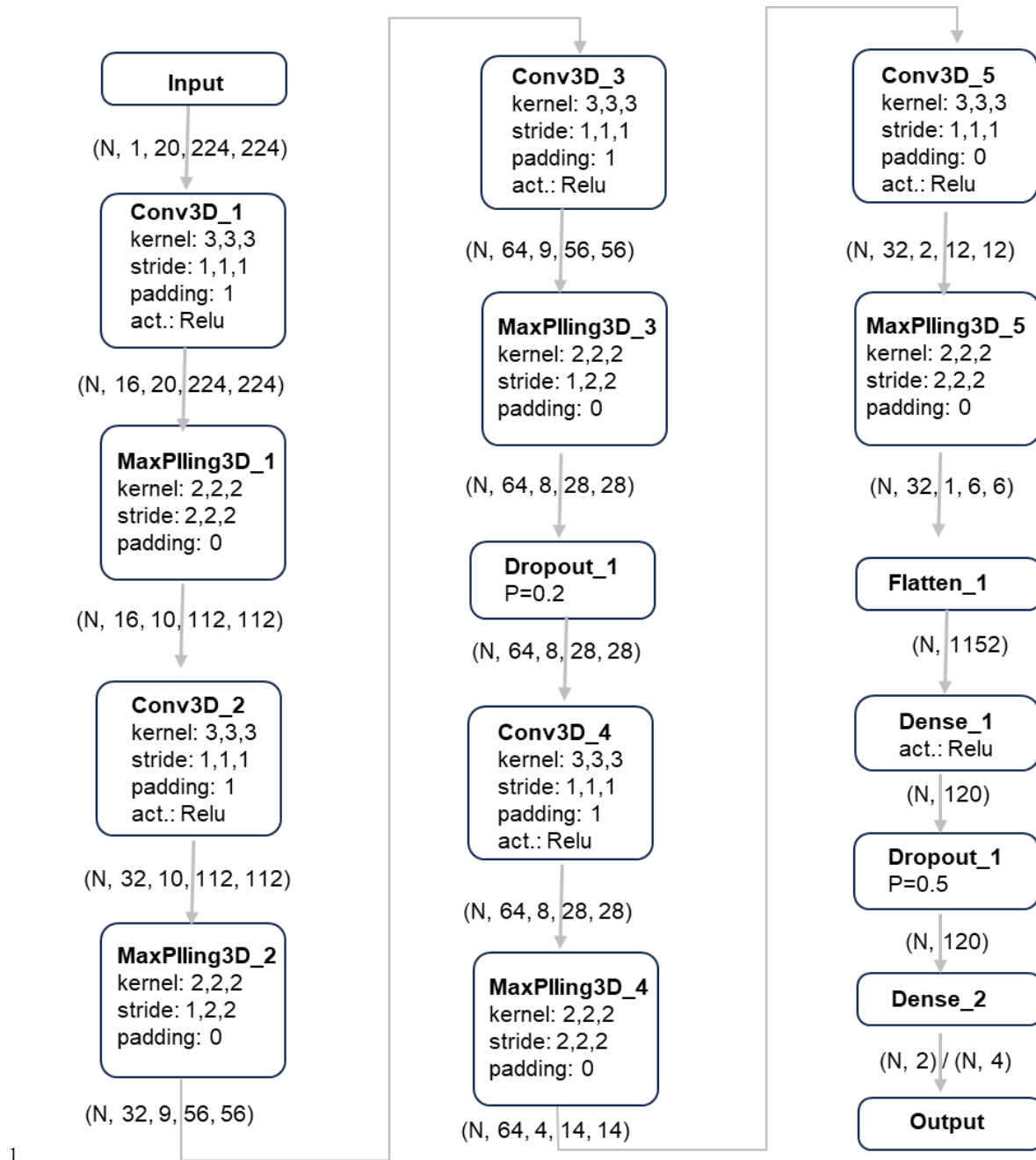
1



2

3 Supplementary Figure S3: Data assimilation phase

4 Data assimilation encompasses a cyclical process that integrates top-down predictions with
 5 bottom-up posterior updates driven by observations. This iterative procedure is conducted in a
 6 sliding time window that retraces H time steps from the current time step t , progressively
 7 moving forward as time advances. During this phase, synaptic weights remain fixed.



Supplementary Figure S4: Network architecture of discriminator (3D convolutional neural network).

1 **References**

- 2 1. Sun, T., X. He, and Z. Li, *Digital twin in healthcare: Recent updates and challenges*.
3 DIGITAL HEALTH, 2023. **9**: p. 20552076221149651.
- 4 2. Corral-Acero, J., et al., *The 'Digital Twin' to enable the vision of precision cardiology*.
5 European Heart Journal, 2020. **41**(48): p. 4556-4564.
- 6 3. Laubenbacher, R., J.P. Sluka, and J.A. Glazier, *Using digital twins in viral infection*.
7 Science, 2021. **371**(6534): p. 1105-1106.
- 8 4. *Computational Psychiatry: New Perspectives on Mental Illness*, ed. A.D. Redish and J.A.
9 Gordon. 2016: The MIT Press.
- 10 5. Markram, H., et al., *Reconstruction and Simulation of Neocortical Microcircuitry*. Cell,
11 2015. **163**(2): p. 456-492.
- 12 6. Yamaura, H., J. Igarashi, and T. Yamazaki, *Simulation of a Human-Scale Cerebellar*
13 *Network Model on the K Computer*. Frontiers in Neuroinformatics, 2020. **14**.
- 14 7. Friston, K.J., et al., *Computational psychiatry: the brain as a phantastic organ*. The Lancet
15 Psychiatry, 2014. **1**(2): p. 148-158.
- 16 8. Otto, A.R., et al., *Working-memory capacity protects model-based learning from stress*.
17 Proceedings of the National Academy of Sciences, 2013. **110**(52): p. 20941-20946.
- 18 9. Ritter, P., et al., *The virtual brain integrates computational modeling and multimodal*
19 *neuroimaging*. Brain Connect, 2013. **3**(2): p. 121-45.
- 20 10. Escuain-Poole, L., J. Garcia-Ojalvo, and A.J. Pons, *Extracranial Estimation of Neural Mass*
21 *Model Parameters Using the Unscented Kalman Filter*. Frontiers in Applied Mathematics
22 and Statistics, 2018. **4**.
- 23 11. Deco, G., et al., *Resting-state functional connectivity emerges from structurally and*
24 *dynamically shaped slow linear fluctuations*. The Journal of neuroscience : the official
25 journal of the Society for Neuroscience, 2013. **33**(27): p. 11239-11252.
- 26 12. Ritter, P., et al., *The virtual brain integrates computational modeling and multimodal*
27 *neuroimaging*. Brain connectivity, 2013. **3**(2): p. 121-145.
- 28 13. Iravani, B., et al., *Whole-brain modelling of resting state fMRI differentiates ADHD*
29 *subtypes and facilitates stratified neuro-stimulation therapy*. NeuroImage, 2021. **231**: p.
30 117844.
- 31 14. Nagasaka, Y., K. Shimoda, and N. Fujii, *Multidimensional Recording (MDR) and Data*
32 *Sharing: An Ecological Open Research and Educational Platform for Neuroscience*. PLOS
33 ONE, 2011. **6**(7): p. e22561.
- 34 15. Yanagawa, T., et al., *Large-Scale Information Flow in Conscious and Unconscious States:*
35 *an ECoG Study in Monkeys*. PLOS ONE, 2013. **8**(11): p. e80845.
- 36 16. Ahmadi, A. and J. Tani. *Bridging the gap between probabilistic and deterministic models:*

- 1 *a simulation study on a variational bayes predictive coding recurrent neural network model.*
2 in *International conference on neural information processing*. 2017. Springer.
- 3 17. Ahmadi, A. and J. Tani, *A novel predictive-coding-inspired variational RNN model for*
4 *online prediction and recognition*. *Neural computation*, 2019. **31**(11): p. 2025-2074.
- 5 18. Idei, H., et al., *Emergence of sensory attenuation based upon the free-energy principle*.
6 *Scientific reports*, 2022. **12**(1): p. 1-15.
- 7 19. Yamashita, Y. and J. Tani, *Emergence of Functional Hierarchy in a Multiple Timescale*
8 *Neural Network Model: A Humanoid Robot Experiment*. *PLOS Computational Biology*,
9 2008. **4**(11): p. e1000220.
- 10 20. Ji, S., et al., *3D Convolutional Neural Networks for Human Action Recognition*. *IEEE*
11 *Transactions on Pattern Analysis and Machine Intelligence*, 2013. **35**(1): p. 221-231.
- 12 21. Jansen, B.H. and V.G. Rit, *Electroencephalogram and visual evoked potential generation*
13 *in a mathematical model of coupled cortical columns*. *Biological Cybernetics*, 1995. **73**(4):
14 p. 357-366.
- 15 22. Yokoyama, H. and K. Kitajo, *A data assimilation method to track excitation-inhibition*
16 *balance change using scalp EEG*. *Communications Engineering*, 2023. **2**(1): p. 92.
- 17 23. Sanz-Leon, P., et al., *Mathematical framework for large-scale brain network modeling in*
18 *The Virtual Brain*. *NeuroImage*, 2015. **111**: p. 385-430.
- 19 24. Deco, G., et al., *How local excitation-inhibition ratio impacts the whole brain dynamics*.
20 *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 2014.
21 **34**(23): p. 7886-7898.
- 22 25. Rivolta, D., et al., *Ketamine Dysregulates the Amplitude and Connectivity of High-*
23 *Frequency Oscillations in Cortical-Subcortical Networks in Humans: Evidence From*
24 *Resting-State Magnetoencephalography-Recordings*. *Schizophrenia Bulletin*, 2015.
25 **41**(5): p. 1105-1114.
- 26 26. McMillan, R. and S.D. Muthukumaraswamy, *The neurophysiology of ketamine: an*
27 *integrative review*. *Reviews in the Neurosciences*, 2020. **31**(5): p. 457-503.
- 28 27. Lee, U., et al., *Disruption of frontal-parietal communication by ketamine, propofol, and*
29 *sevoflurane*. *Anesthesiology*, 2013. **118**(6): p. 1264-75.
- 30 28. Schroeder, K.E., et al., *Disruption of corticocortical information transfer during ketamine*
31 *anesthesia in the primate brain*. *Neuroimage*, 2016. **134**: p. 459-465.
- 32 29. Friston, K., J. Kilner, and L. Harrison, *A free energy principle for the brain*. *Journal of*
33 *Physiology-Paris*, 2006. **100**(1-3): p. 70-87.
- 34 30. Friston, K., *The free-energy principle: a unified brain theory?* *Nature reviews*
35 *neuroscience*, 2010. **11**(2): p. 127-138.
- 36 31. Idei, H., et al., *A Neurorobotics Simulation of Autistic Behavior Induced by Unusual*

- 1 *Sensory Precision*. Computational Psychiatry, 2018: p. 1-19.
- 2 32. Idei, H., et al., *Homogeneous Intrinsic Neuronal Excitability Induces Overfitting to*
3 *Sensory Noise: A Robot Model of Neurodevelopmental Disorder*. Front Psychiatry, 2020.
4 **11**: p. 762.
- 5 33. Takahashi, Y., et al., *Neural network modeling of altered facial expression recognition in*
6 *autism spectrum disorders based on predictive processing framework*. Scientific Reports,
7 2021. **11**(1): p. 14684.
- 8 34. Takahashi, Y., et al., *Interaction between Functional Connectivity and Neural Excitability*
9 *in Autism: A Novel Framework for Computational Modeling and Application to Biological*
10 *Data*. Computational Psychiatry, 2023.
- 11 35. Soda, T., et al., *Simulating developmental diversity: Impact of neural stochasticity on*
12 *atypical flexibility and hierarchy*. Frontiers in Psychiatry, 2023. **14**.
- 13 36. Komatsu, M., et al., *An artificial network model for estimating the network structure*
14 *underlying partially observed neuronal signals*. Neuroscience Research, 2014. **81-82**: p.
15 69-77.
- 16 37. Komatsu, M., K. Takaura, and N. Fujii, *Mismatch negativity in common marmosets:*
17 *Whole-cortical recordings with multi-channel electrocorticograms*. Scientific Reports,
18 2015. **5**(1): p. 15006.
- 19 38. Friston, K. and S. Kiebel, *Predictive coding under the free-energy principle*. Philosophical
20 Transactions of the Royal Society B: Biological Sciences, 2009. **364**(1521): p. 1211-1221.
- 21 39. Liu, L., et al., *On the variance of the adaptive learning rate and beyond*. arXiv preprint
22 arXiv:1908.03265, 2019.
- 23 40. Craik, A., Y. He, and J.L. Contreras-Vidal, *Deep learning for electroencephalogram (EEG)*
24 *classification tasks: a review*. Journal of Neural Engineering, 2019. **16**(3): p. 031001.
- 25 41. Rousseeuw, P.J., *Silhouettes: A graphical aid to the interpretation and validation of cluster*
26 *analysis*. Journal of Computational and Applied Mathematics, 1987. **20**: p. 53-65.

27