Research Plan

The research delivers 1) machine-learning models that forecast the target state given time moment with the target state previous time and source state given time, 2) shows the dependency of the target states of the source states for selected time segments, and 3) analyses model errors on several datasets. *Below the "source" mean EEG, ECoG, and iEEG-like signals, in other words, the brain signals taken from the surface of high informative frequency (<170Hz). The "target" means IMU or video signals, of the physical motion of low informative frequency (<10Hz).* The following tasks are planned with respect risks of research and development:

1. Select two datasets, and a test one. The alternatives are a) Mind in motion young adults walking over uneven terrain, b) Real-world table tennis, and c) a generated synthetic dataset to comply with the MLOps requirements.
2. Prepare data for modeling: select an NME-compatible format, convert to this format, and clean. Generate a dataset.
3. Select alternative Sequence to Sequence, Space-State Models: two or three models like RNN-LSTM models with cross-attention mechanism. Test also self-attention for both source and target signals.
4. Code the models and test on the synthetic data.
5. Construct and run the computational experiment and error analysis.
6. Estimate and report forecasting accuracy.
7. Simplify models to make a baseline for accuracy and complexity analysis. Select one of the Canonical Correlation Analysis libraries, if possible.
8. Run the CCA model to determine the optimal dimensionality of the source, target, and latent space. Analyze the geometry (changes in covariance matrices over time and parameters) of the latent space.
9. List the hyperparameters of the created models to perform the Ganger casualty test on the signal dependency.
10. Code the Granger test and run the experiment to estimate the casualty over time. Label the time with an expert markup and analyze the results of the test.
11. Boost the forecast accuracy and casualty by creating a pipeline for data processing and model selection. The following models might be included in the pipeline to process the data. For the source: a) spectrogram and scalogram representations, b) bandpass filtering, failure detection, drift detection, muscle movement detection, and Riemannian state space representation. For the target: a) failure and drift detection, b) Singular Spectrum Analysis decomposition, c) Physics-informed neural networks approximation.
12. Create a decomposition procedure to find the first principal or independent components in the source state space.

13. Select a library to perform the Convergent Cross Mapping casualty test or code it.
14. Reconstruct variants of the latent space for different groups of principal or independent components of the source state. Code the PCA/ICA decomposition of the source state. This space could be the same for the source and the target (if we are lucky with its dimensionality). The other way, there are two state spaces: constructed for the source and the target separately following a common timeline.
15. Elaborate the CCM casualty criterion (set a schedule to sample the neighborhood in the target and reconstruct it in the source space). Code and run the computational experiment.
16. Compare the CCM results with the Granger test results. If possible, code the Debiased Orthogonal Machine Learning casualty test to compare the results.
17. Select the necessary components of decomposition of the source space. Construct an inverse model, to set source-related quality criteria that describe physical motions from the target.
18. Propose and code variants of the motion quality criteria.
19. Code and run the computational experiment on the data with the expert-marked timeline.
20. Compare and analyze a) the quality criteria, b) the forecasting accuracy and the Granger test, c) CCM casualty tests, and d) labels from the expert-marked timeline. Interpret the results according to the expert estimations.
21. Make the cross-participant and the session-to-session analysis. To do this step we have to rerun tasks 5–20 from this list and estimate the changes in the model structures, model parameters, and distributions of obtained results. If there will be significant changes these tasks shall be remade. This is one of the two main risks of the project. The second one is obvious: there is no casualty between the source and target detected with the selected models under the given conditions of the signal measurements.
22. Report the results of cross-participant analysis. List the model-pretraining items and necessary tasks for the pre-training calibration.
23. Make the system of models, tests, and quality criteria. Gather the unit tests in suits, and check the MLOps compliance.
24. Create cases to run field experiments with measurement analysis and deploy.

This is a preliminary plan. Plenty of less significant, but time-consuming tasks on coding, testing, system organizing, and reporting were put aside.

Technical Objectives

To elaborate reporting criteria that show the statistical significance of the presence of target signals in the source signals, in other words, the dependency of the target signal on the source.
We assume that the target is described by a unique manifold in the spate space, but the source contains a combination (most likely, linear) of several sources. So we have to find a submanifold to establish the mentioned dependency. The dependency on the casualty model could differ from the forecasting model. The last one forecasts the state and the signal values at the given time sample depending on the current source state and previous target state. We include a long-term state or a context in the forecasting model. It is a participant description and a session description.
The timeline. The event-related potential signals are easier to detect and forecast. In our ERP, there is a presence of stimulus type in the EEG signal. The ERP is the brain-induced signal
The submanifolds. The EEG is acquired from the surface of the skull. We assume that there are several main sources of signals inside the skull [proof]. Each source corresponds to its own manifold.
The intracranial space (reconstructed from the EEG, could be checked with the fMRI signal). Several areas of the brain could show an active state. We assume a single area to correspond to the EEG signal.