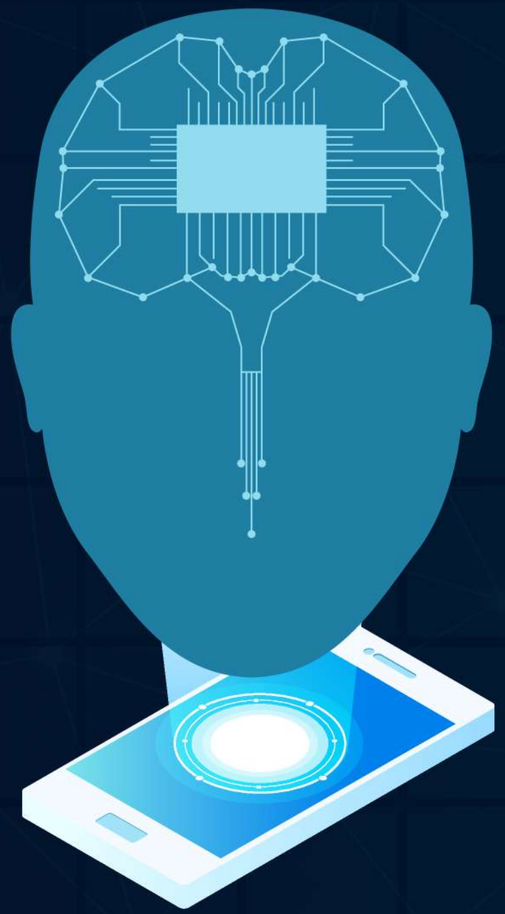


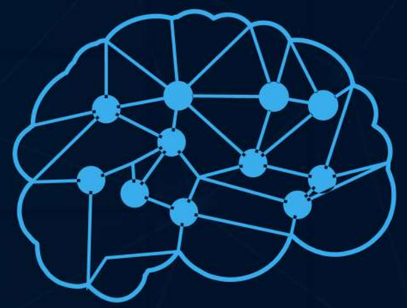
# 50.039 THEORY AND PRACTICE OF DEEP LEARNING

## PROJECT PRESENTATION


PITCH DECK BY :

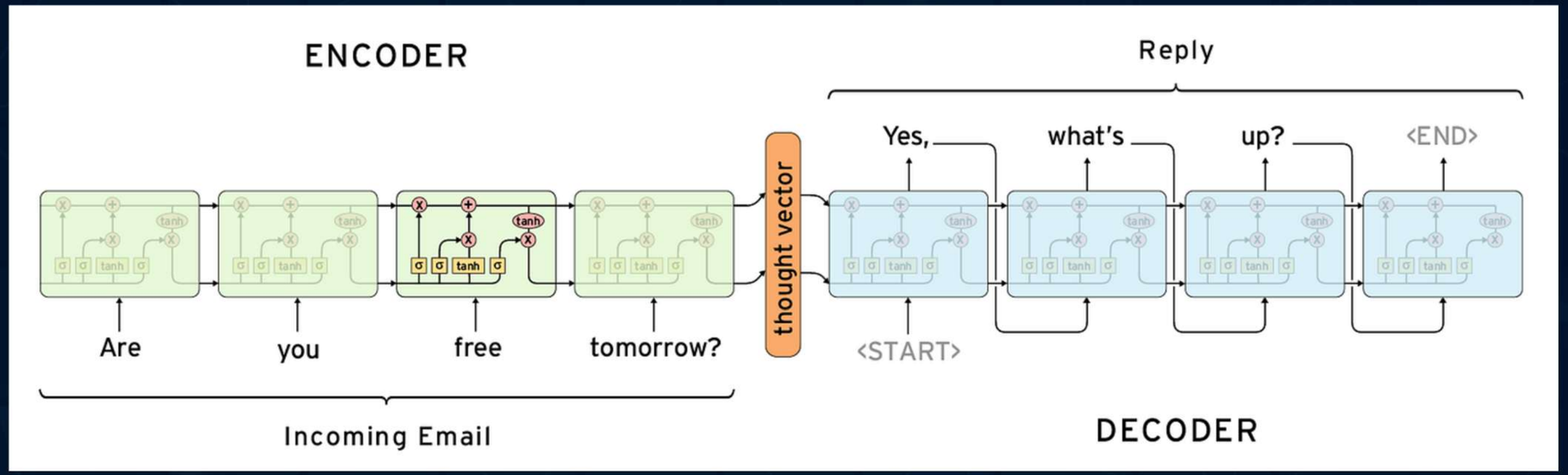
TRAN CONG NAM ANH
OON SHAO REN
BENETTA CHENG JIA WEN

# IMPLEMENTATION PLAN AND STATUS

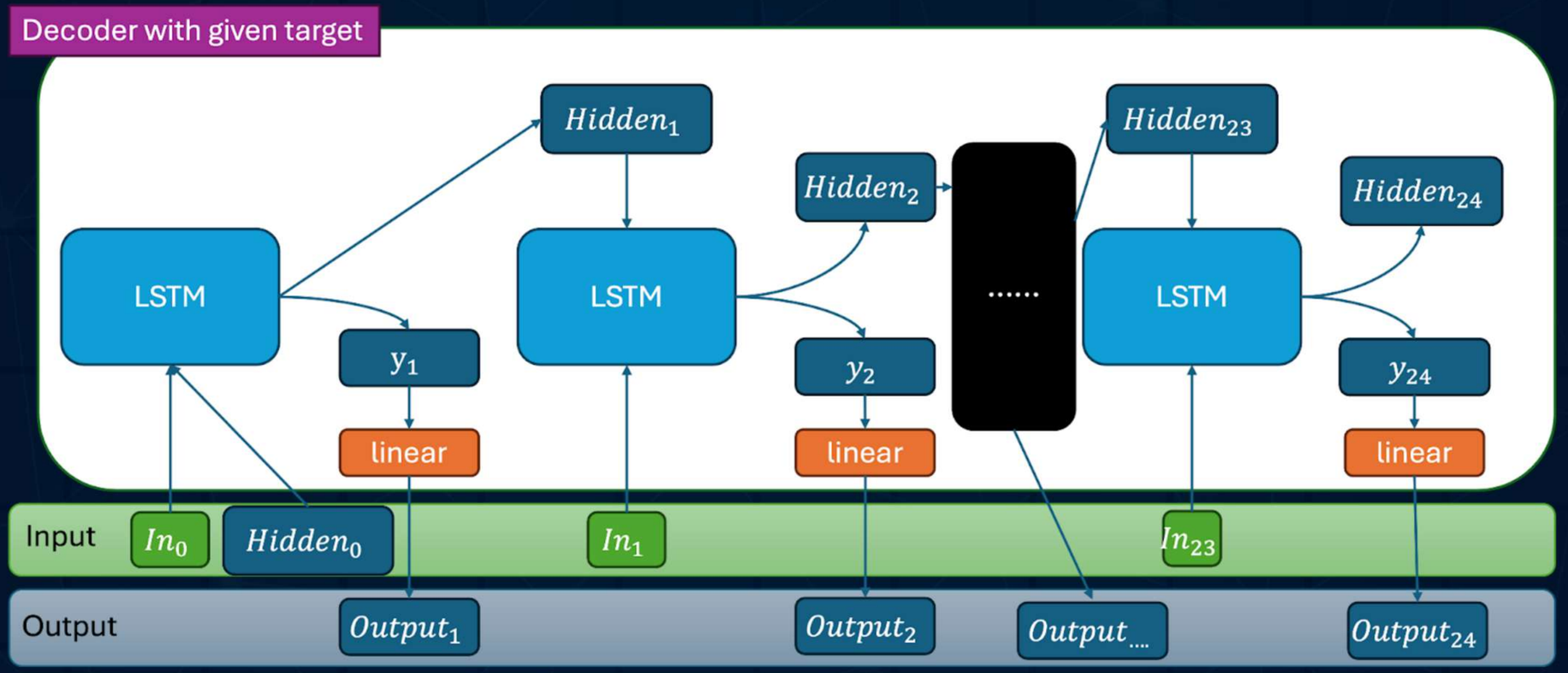


| Models | Scope | Status |
|---|---|---|
| Seq2Seq auto-regressive Recurrent Neural Network using Long Short-Term Memory (LSTM) | Train from scratch | Completed |
| Seq2Seq non-auto-regressive Recurrent Neural Network using Long Short-Term Memory (LSTM) | Train from scratch | Completed |
| Seq2Seq auto-regressive Recurrent Neural Network using Gated Recurrent Unit (GRU) | Train from scratch | Completed |
| Seq2Seq non-auto-regressive Recurrent Neural Network using Gated Recurrent Unit (GRU) | Train from scratch | Completed |
| Recurrent Neural Network with Transformer architecture | Train from scratch | Completed |
| Use/Fine-tune state-of-the-art pre-trained model as the baseline for comparison | Fine tune using Transfer learning | In progress |

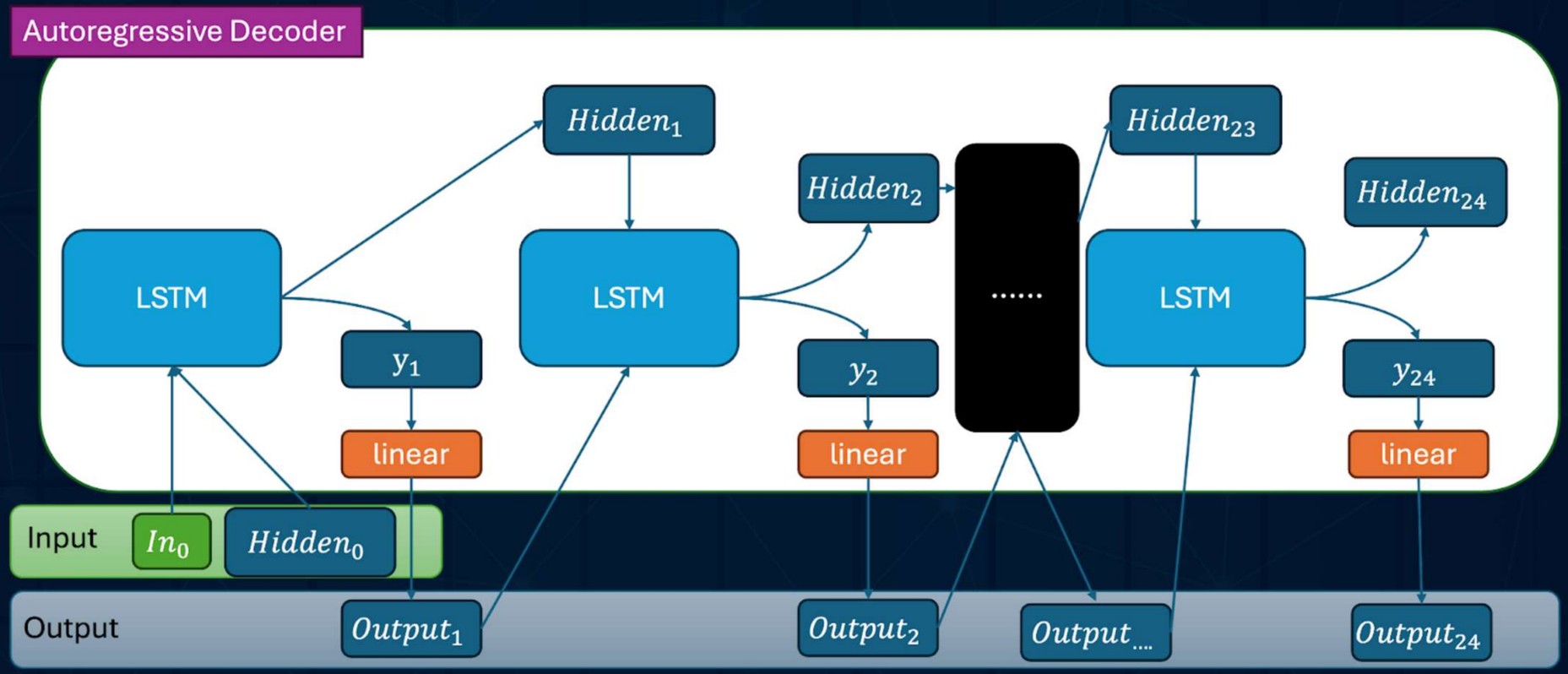
# SEQ2SEQ RNN MODEL USING LSTM



# SEQ2SEQ RNN MODEL USING LSTM



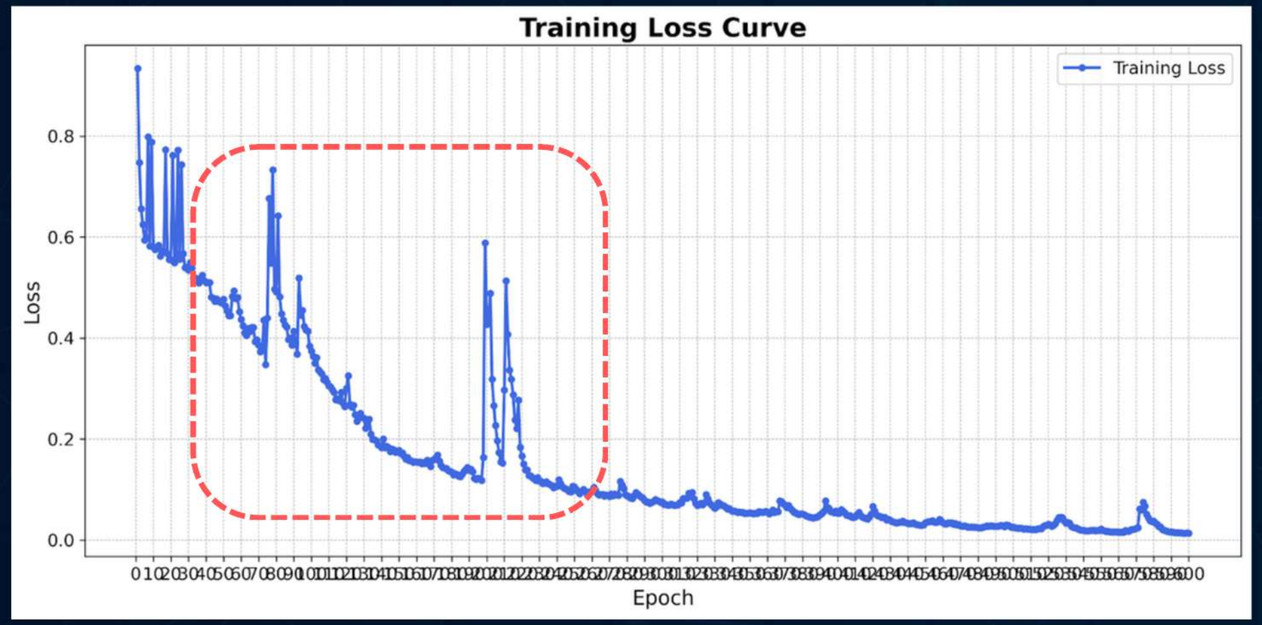
# SEQ2SEQ RNN MODEL USING LSTM



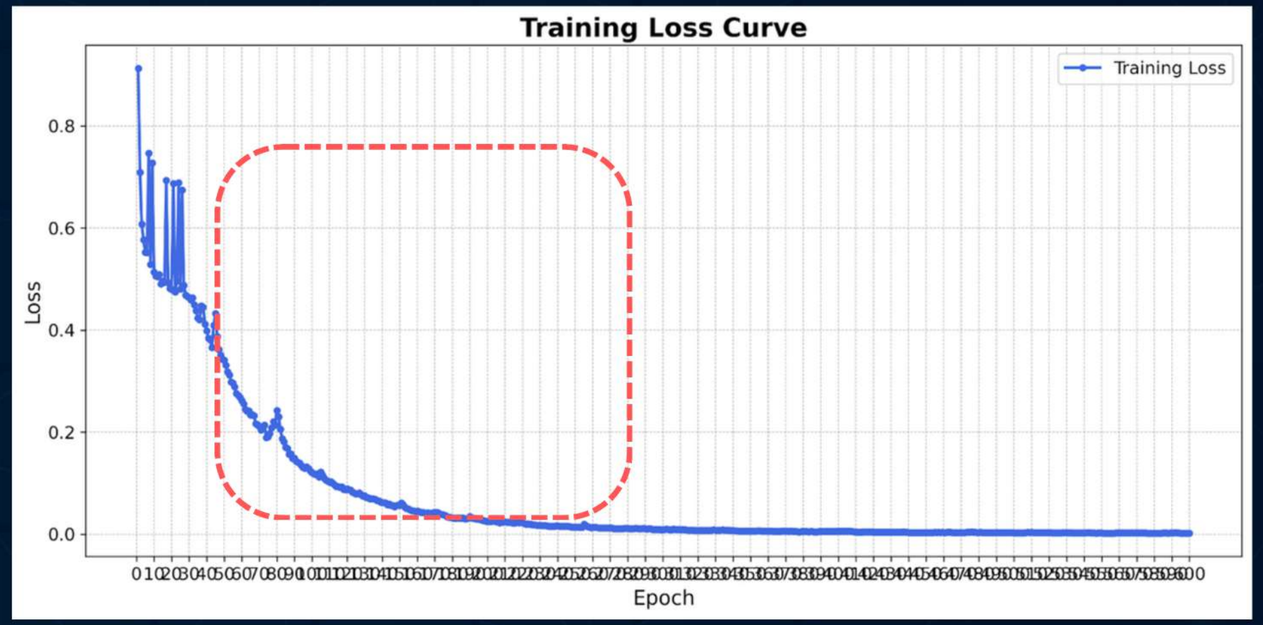
# 1 SEQ2SEQ RNN MODEL USING LSTM

| Variables | Hyperparameter Details | Values |
|---|---|---|
| HIDDEN_SIZE | The dimensions of memory vector, which consist of hidden state and cell state | 100 |
| LEARNING_RATE | The learning rate used in gradient descent formula | 0.001 |
| NUM_EPOCHS | The number of epochs to be used during model training | 600 |
| NUMBER_LAYERS | Number of layers in each LSTM layer in both encoder and decoder | 2 |
| DROPOUT_RATE | The rate for dropout layer | 0.1 |
| N_INPUTS | The number of input sequence length, meaning how many data points the model used as the input | 30 |
| N_OUTPUTS | The number of output sequence length, meaning how many data points in the future that model will generate | 7 |

# 1 SEQ2SEQ RNN MODEL USING LSTM

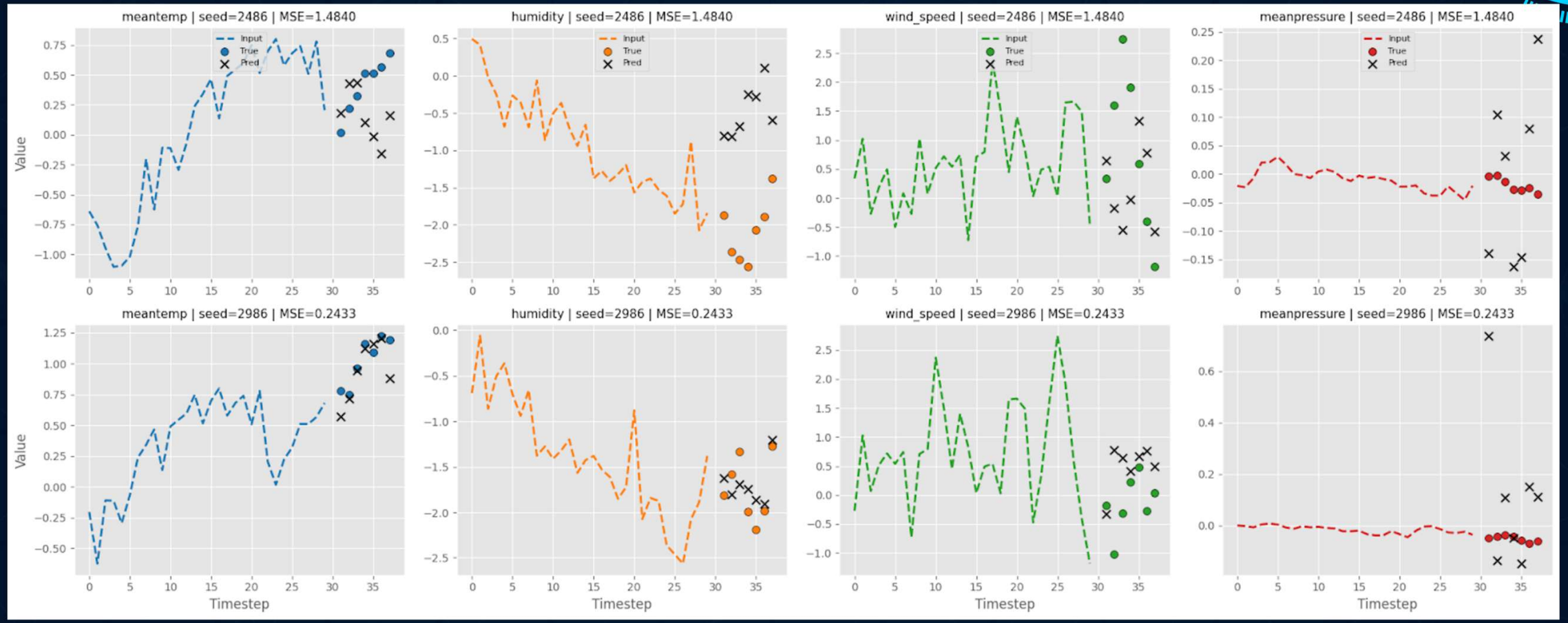



AUTO-REGRESSIVE



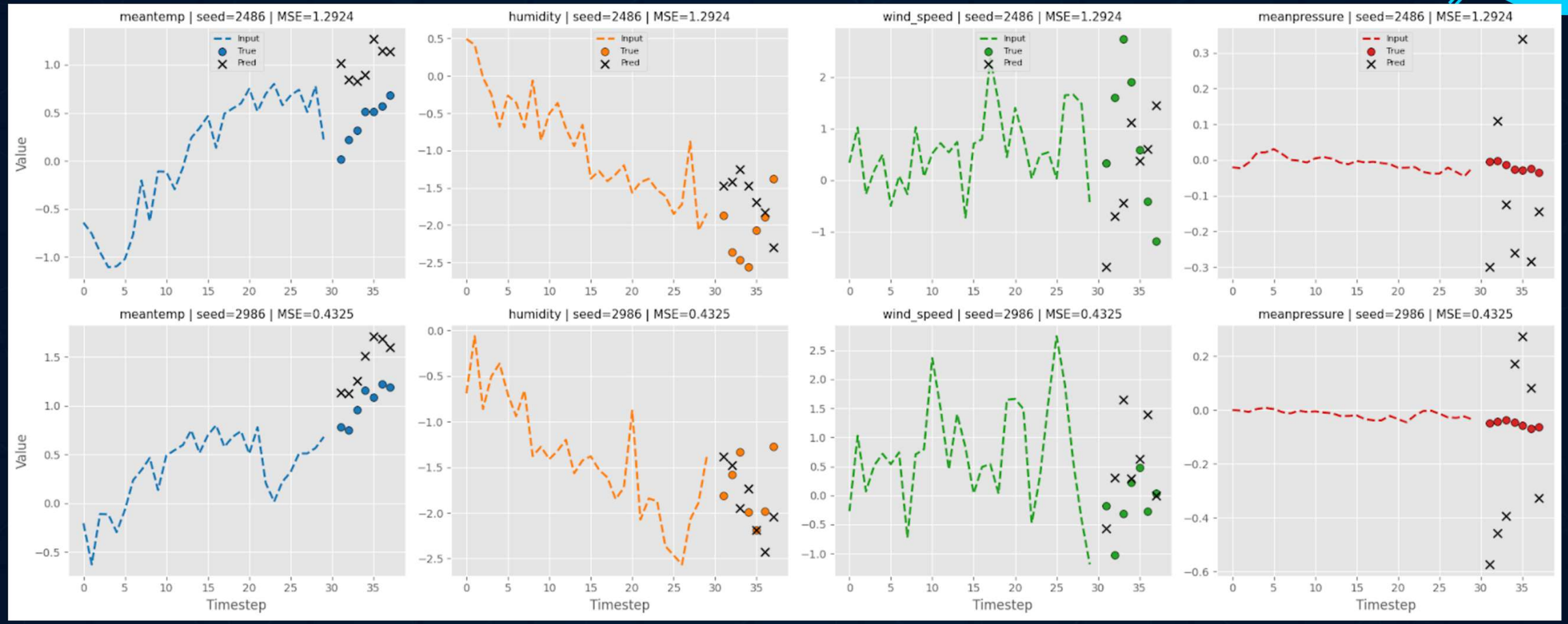

NON AUTO-REGRESSIVE

# 1 SEQ2SEQ RNN MODEL USING LSTM



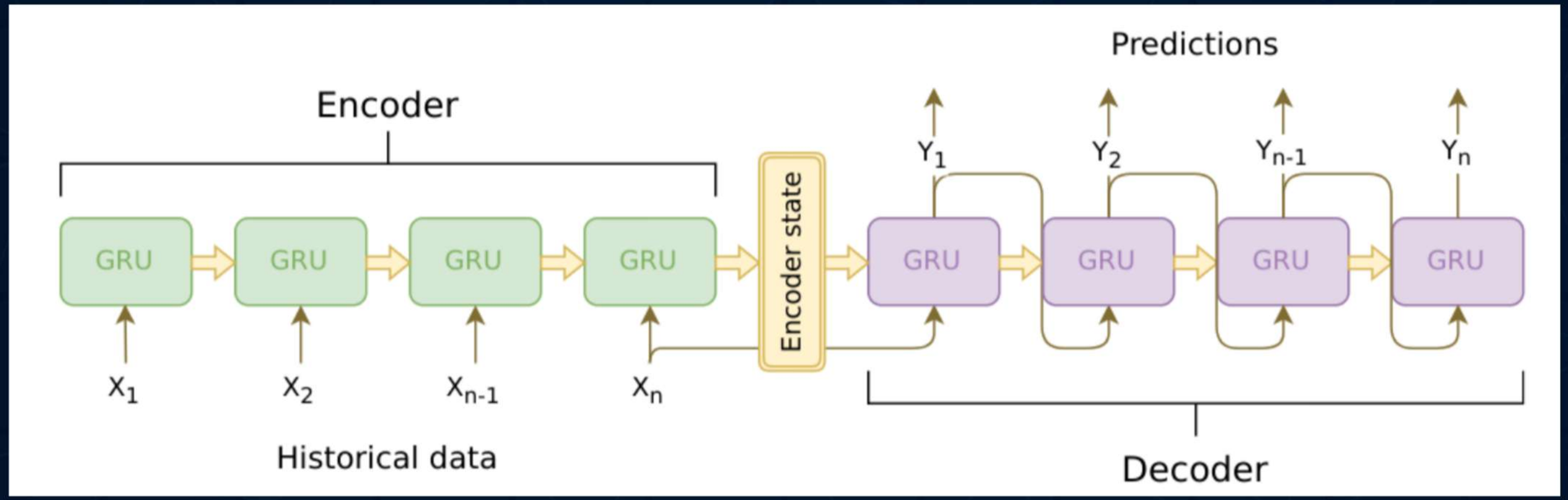

# AUTO-REGRESSIVE MODEL

# SEQ2SEQ RNN MODEL USING LSTM

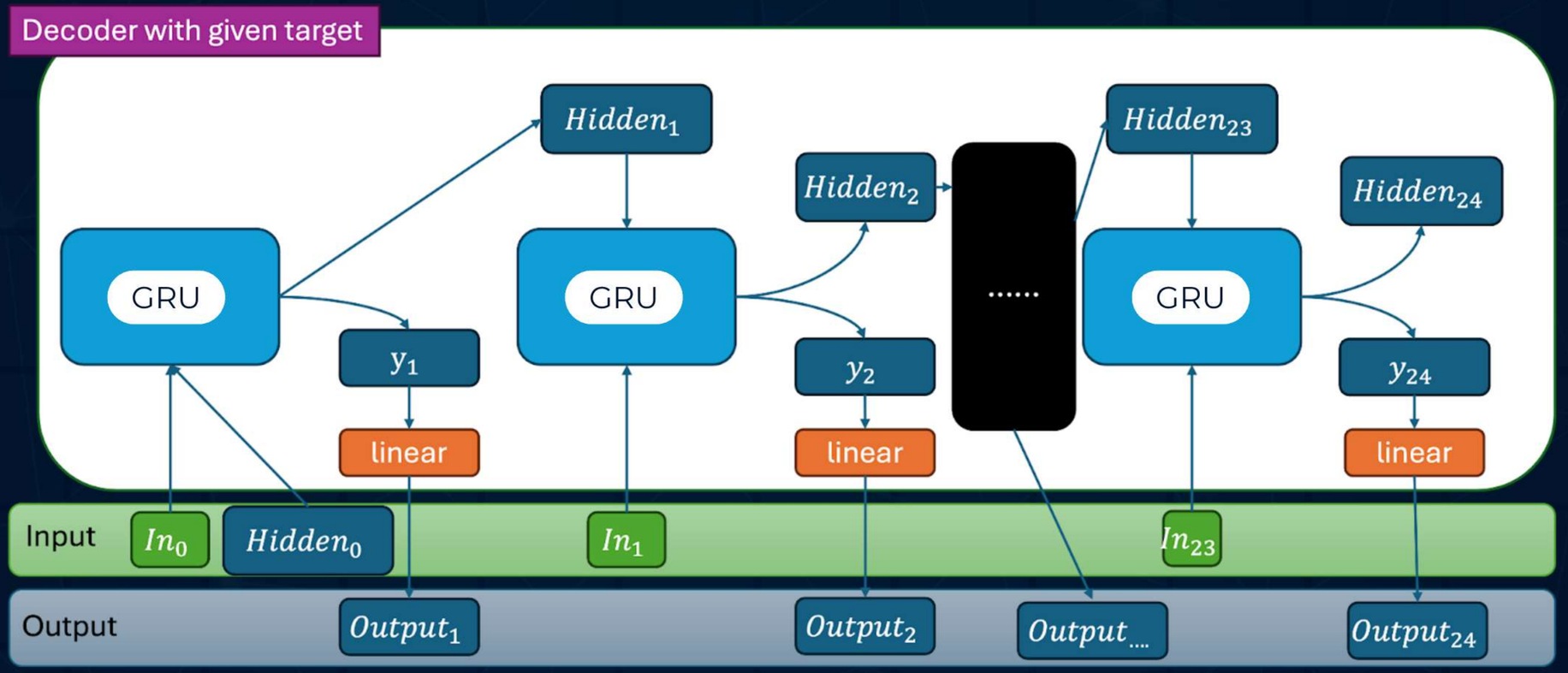



NON AUTO-REGRESSIVE MODEL

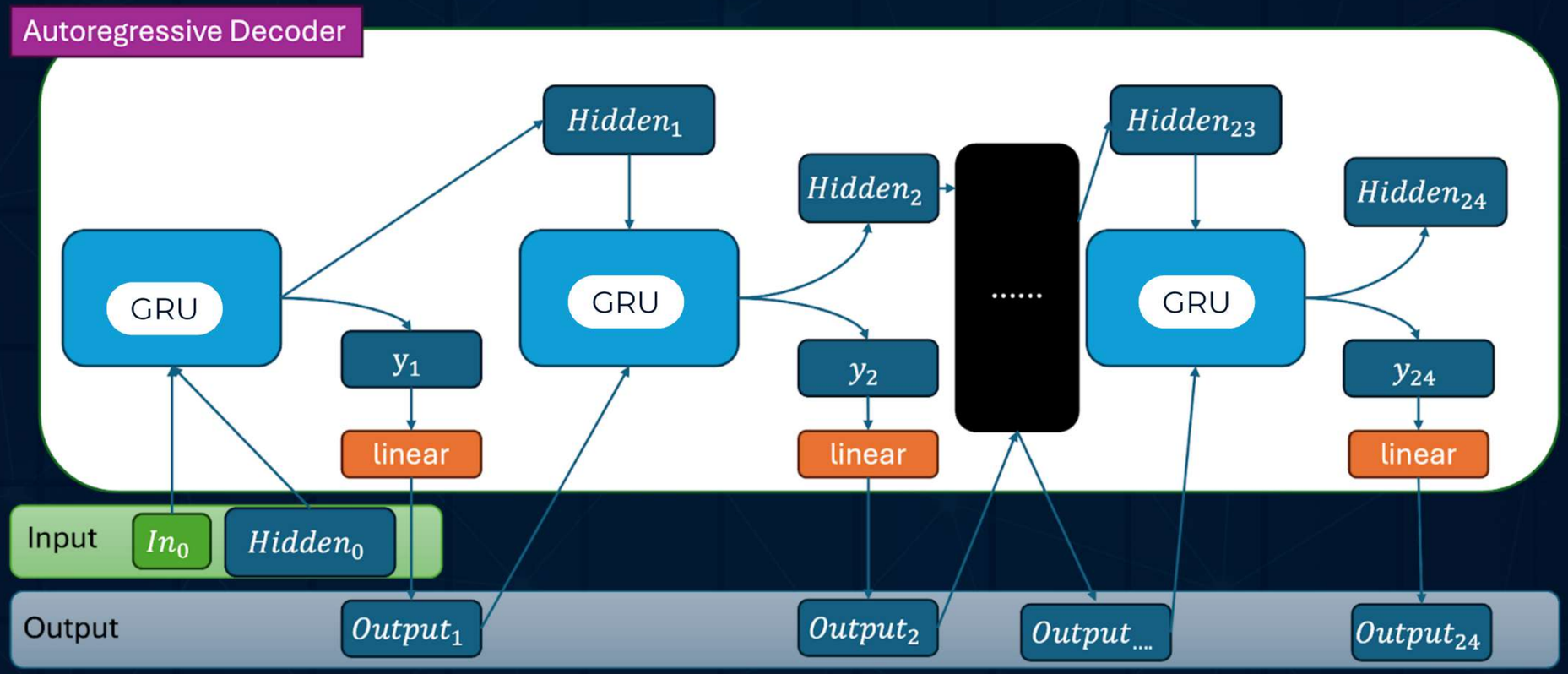
# SEQ2SEQ RNN MODEL USING GRU



# SEQ2SEQ RNN MODEL USING GRU



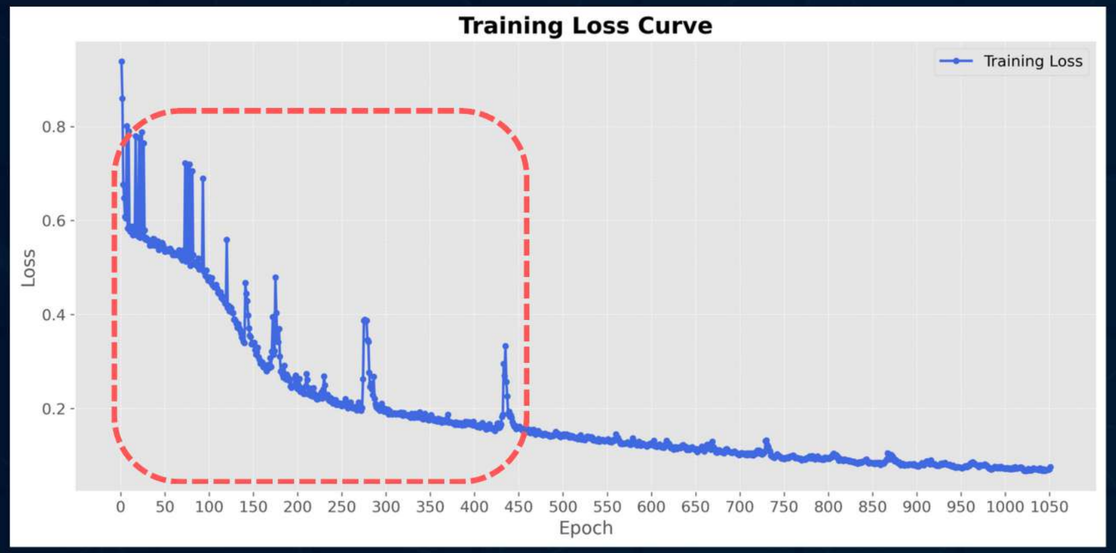
# SEQ2SEQ RNN MODEL USING GRU



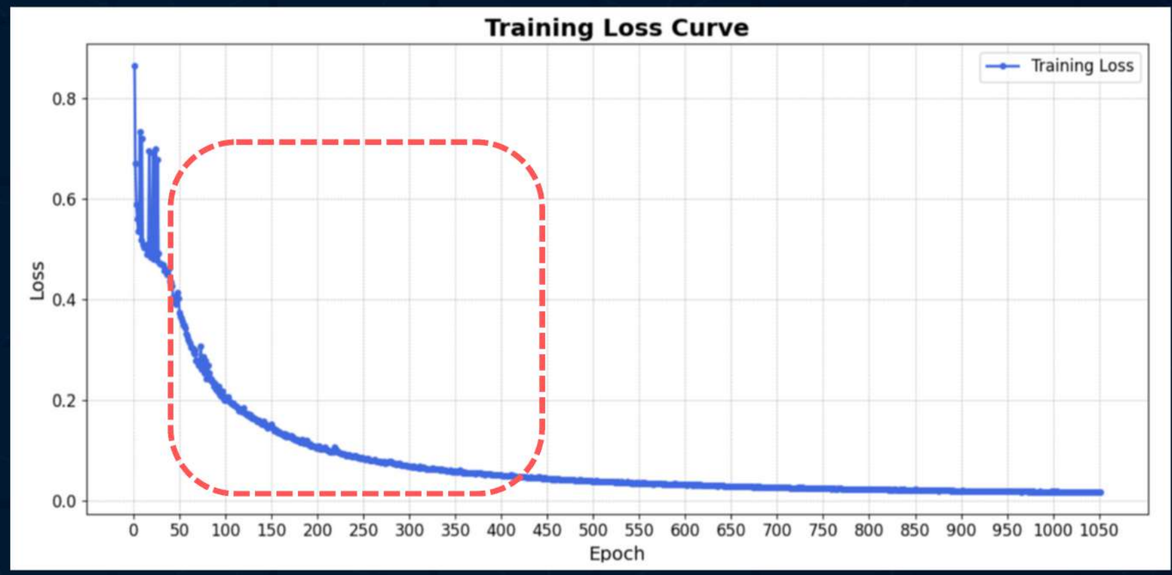
# SEQ2SEQ RNN MODEL USING GRU

| Variables | Hyperparameter Details | Values |
|---|---|---|
| Data dimensions | | |
| N_INPUTS | The number of input sequence length, meaning how many data points the model used as the input | 30 |
| N_OUTPUTS | The number of output sequence length, meaning how many data points in the future that model will generate | 7 |
| Model architecture hyperparameters | | |
| HIDDEN_SIZE | The dimensions of memory vector, which consist of hidden state and cell state | 64 |
| NUMBER_LAYERS | Number of layers in each LSTM layer in both encoder and decoder | 1 |
| DROPOUT_RATE | The rate for dropout layer | 0 |
| Training hyperparameters | | |
| LEARNING_RATE | The learning rate used in gradient descent formula | 0.001 |
| NUM_EPOCHS | The number of epochs to be used during model training | 1051 |

# SEQ2SEQ RNN MODEL USING GRU

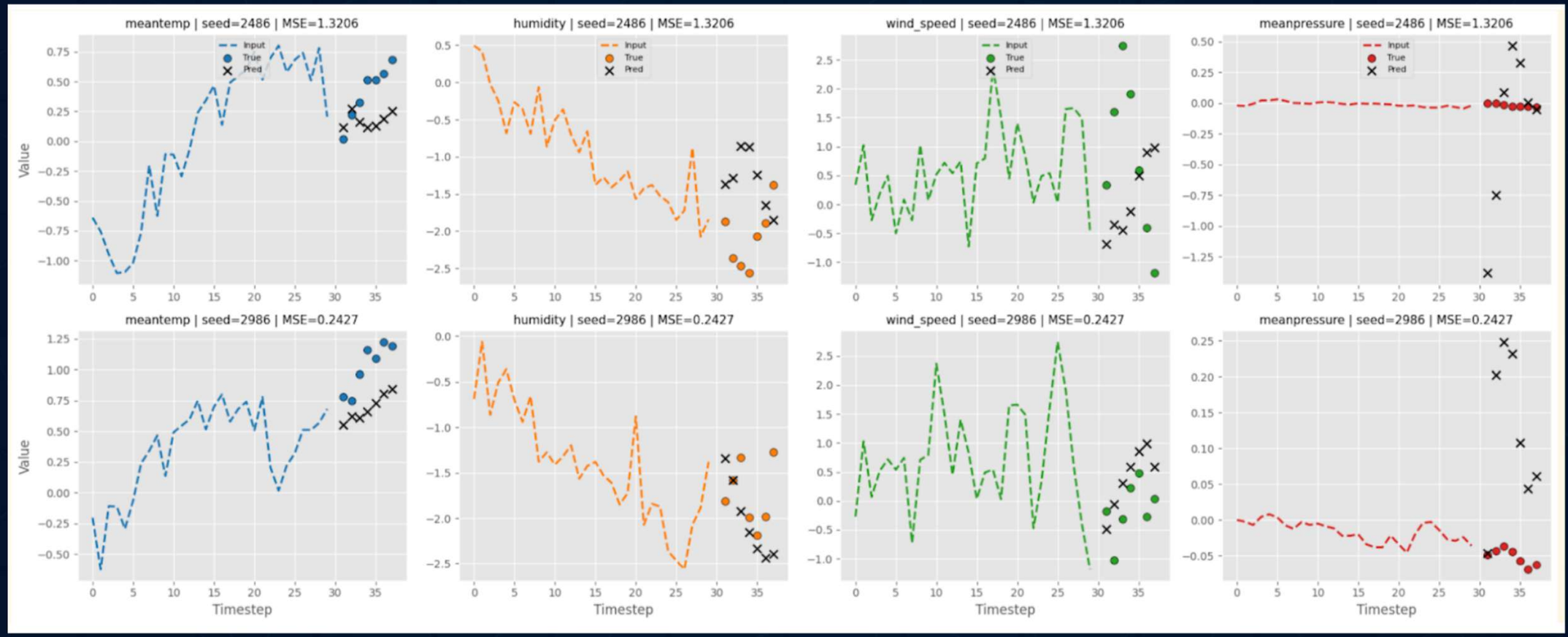



AUTO-REGRESSIVE



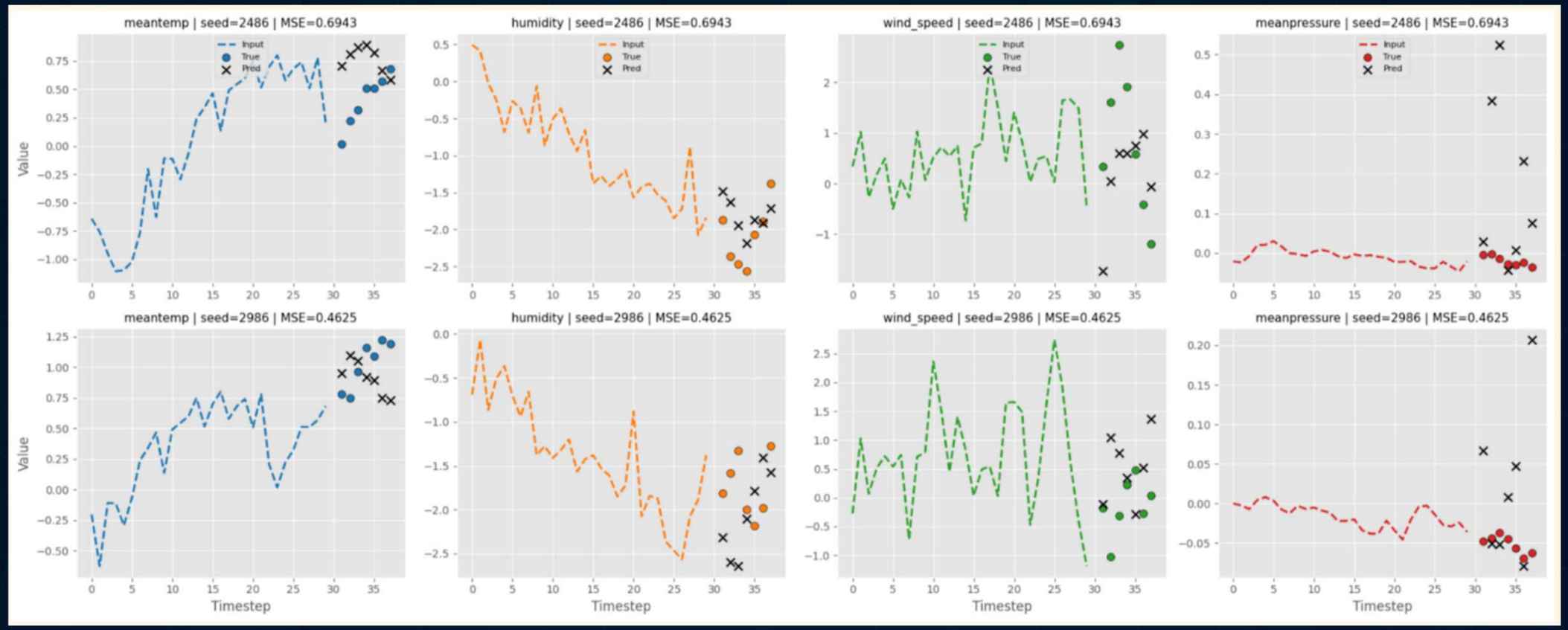

NON AUTO-REGRESSIVE

# SEQ2SEQ RNN MODEL USING GRU




# AUTO-REGRESSIVE MODEL

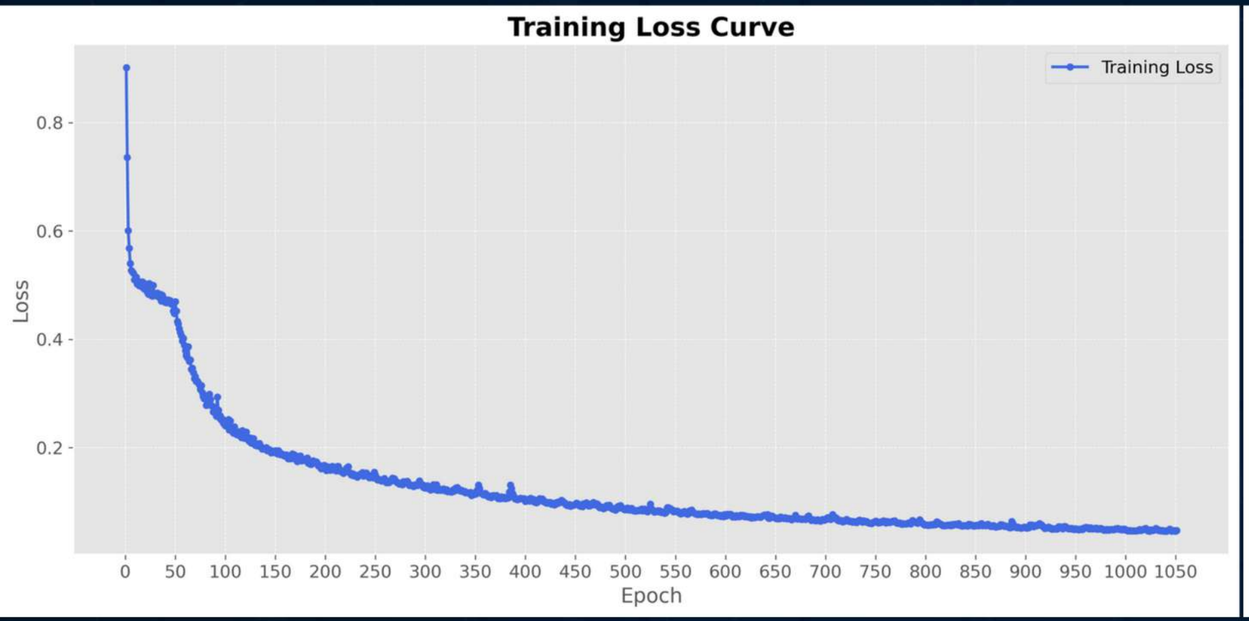
# SEQ2SEQ RNN MODEL USING GRU

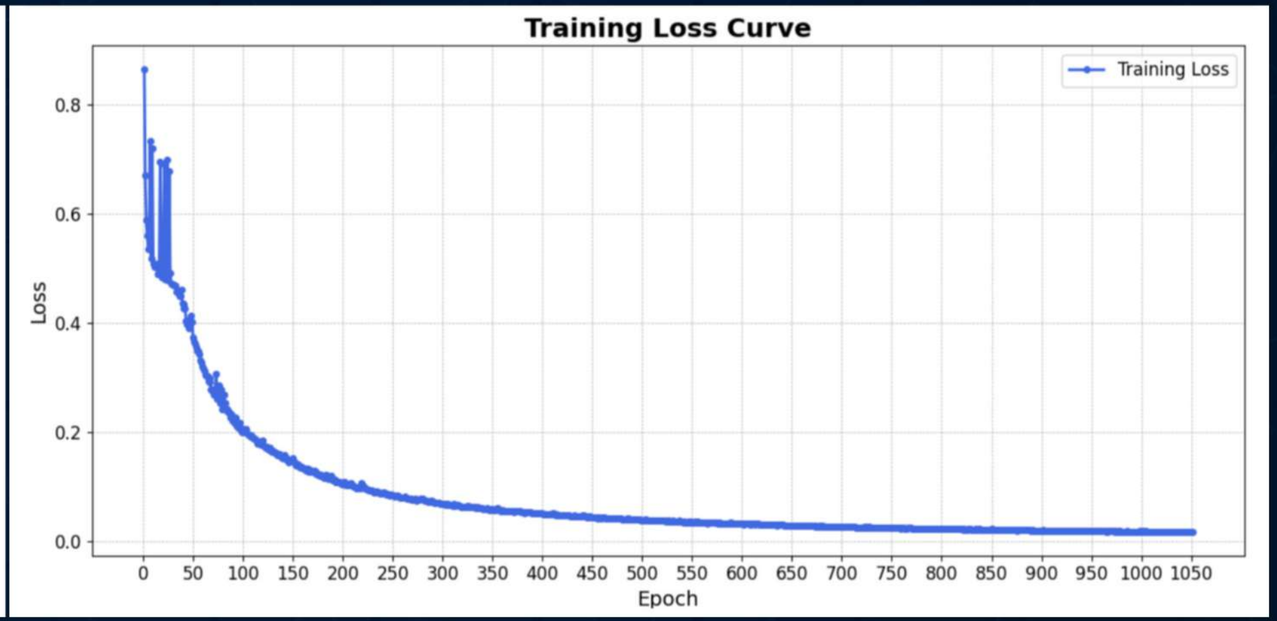



NON AUTO-REGRESSIVE MODEL

# 2

# IMPROVING TRAINING OF MODEL - SEQ2SEQ GRU NON-AUTOREGRESSIVE



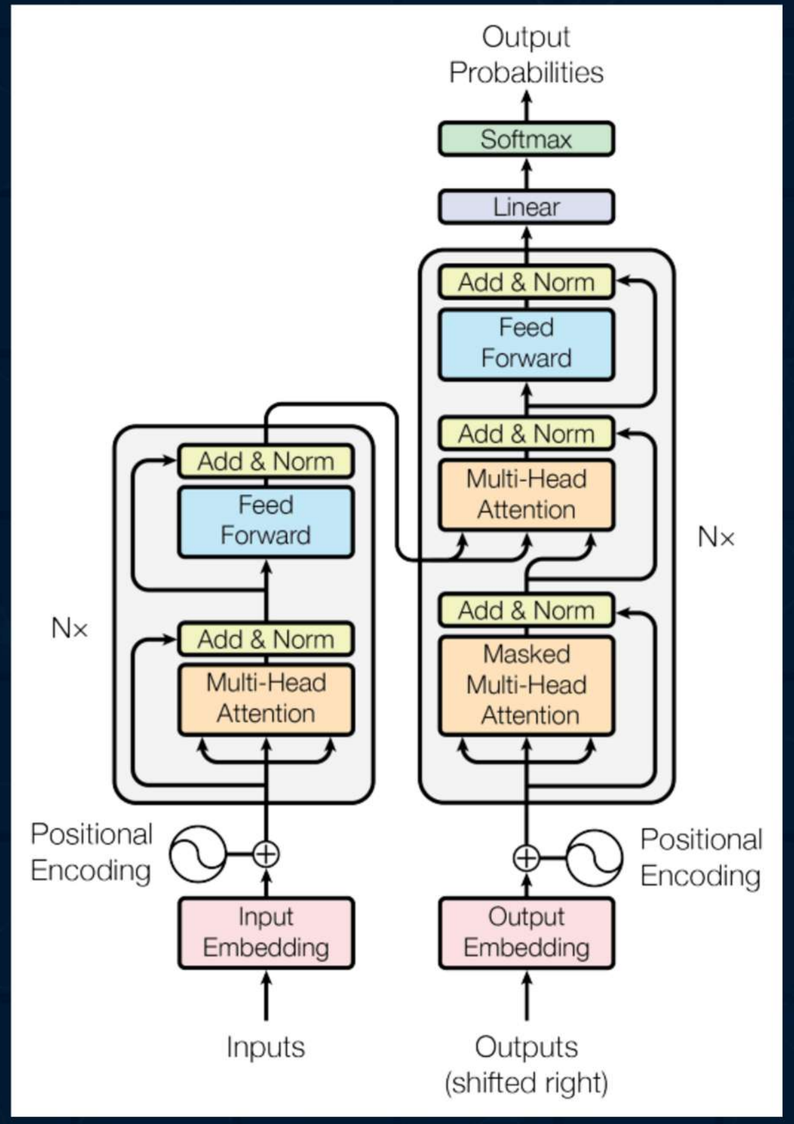

NUM_LAYERS: 1
DROPOUT: 0

NUM_LAYERS: 2
DROPOUT: 0.1

# IMPROVING TRAINING OF MODEL - SEQ2SEQ GRU NON-AUTOREGRESSIVE

| Metric | 1 Layer, No Dropout | 2 Layers, Dropout 0.1 | Improvement |
|---|---|---|---|
| Final Loss (Epoch 1050) | 0.0416 | 0.0173 | 2.4x lower |
| Test MAE | 0.3725 | 0.2317 | ~37.8% lower |
| Training Time | 2m 6s | 4m 52s | 2.3x longer |

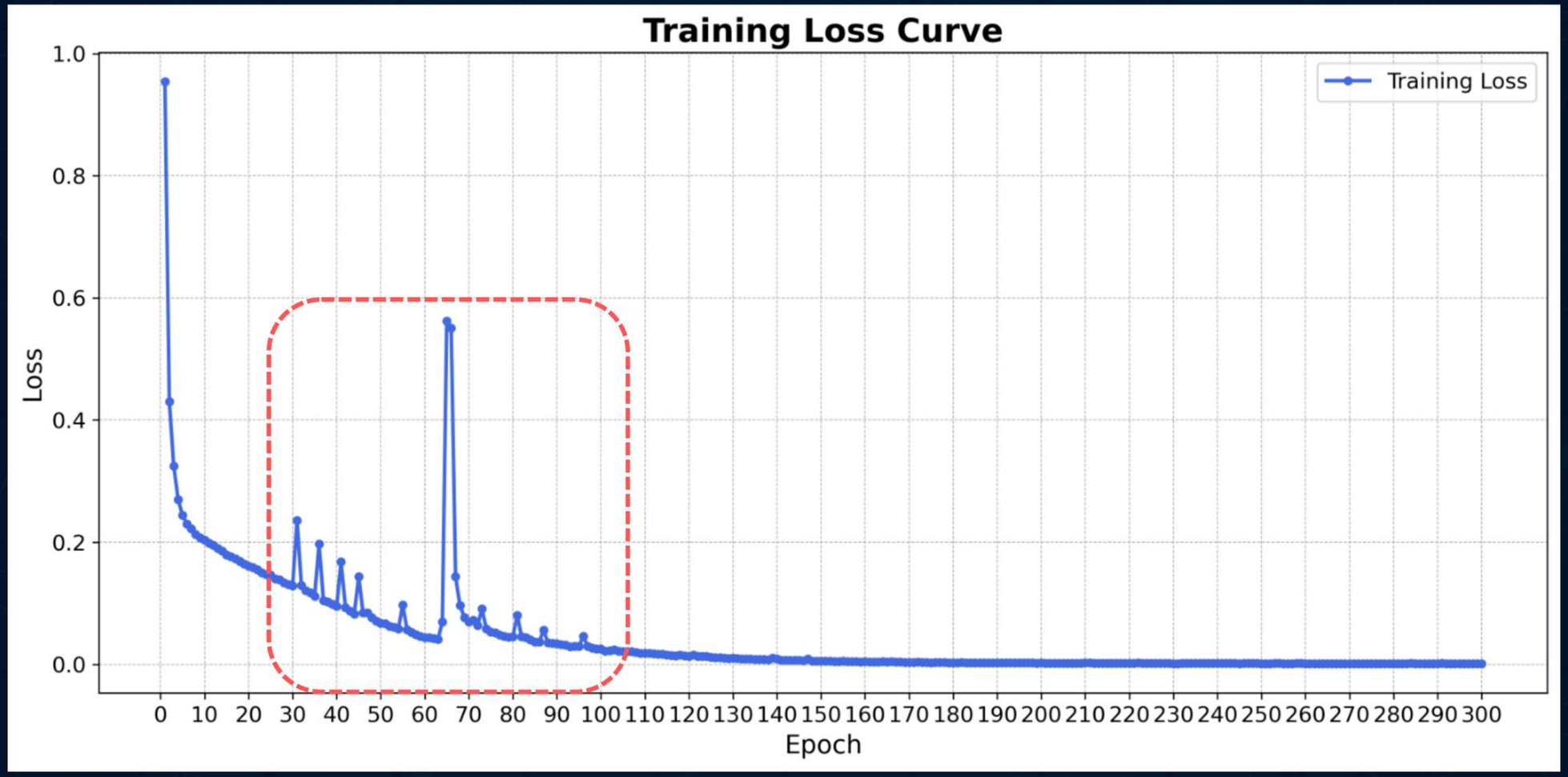
# 3 SEQ2SEQ MODEL USING TRANSFORMER



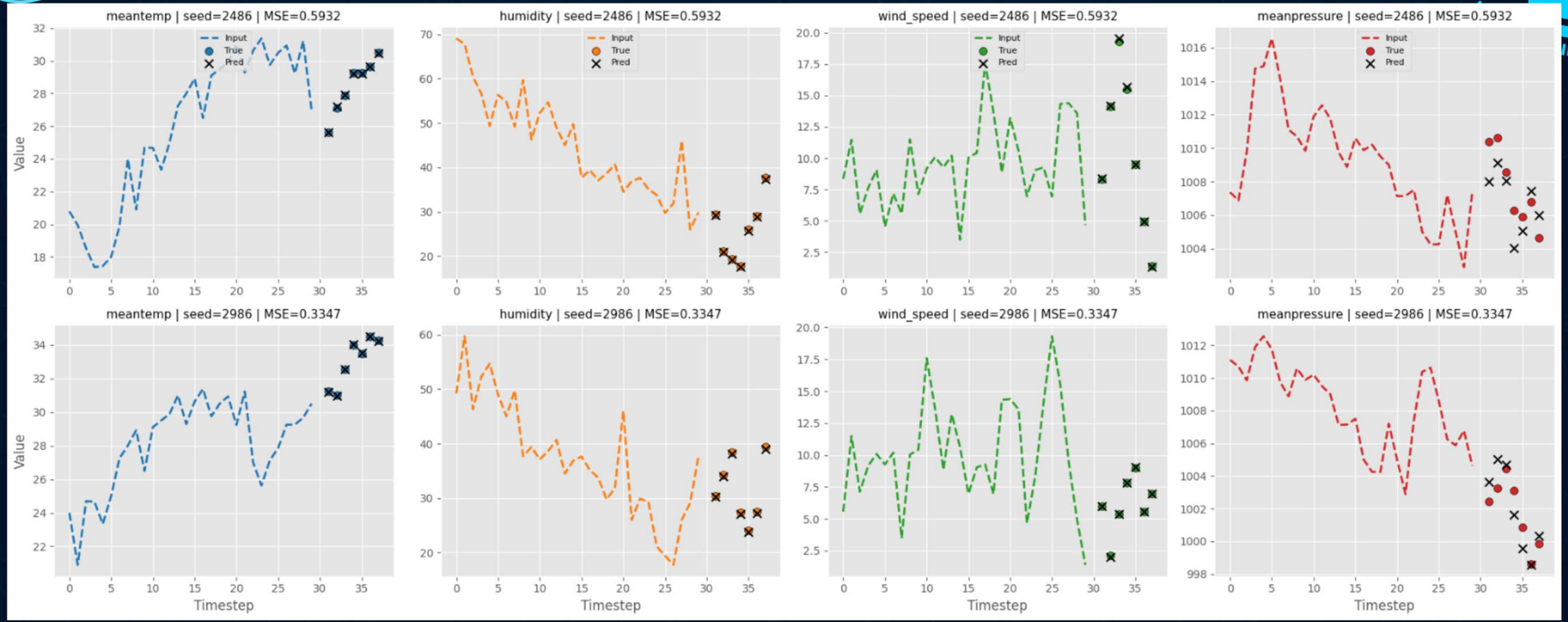
# 3 SEQ2SEQ MODEL USING TRANSFORMER

| Variables | Hyperparameter Details | Values |
|---|---|---|
| Data dimensions | | |
| N_INPUTS | The number of input sequence length, meaning how many data points the model used as the input | 30 |
| N_OUTPUTS | The number of output sequence length, meaning how many data points in the future that model will generate | 7 |
| Model architecture hyperparameters | | |
| HIDDEN_SIZE | The dimensions of memory vector, which consist of hidden state and cell state | 64 |
| NHEAD | The number of attention heads | 4 |
| NUMBER_LAYERS | Number of encode and decoder layers | 2 |
| DROPOUT_RATE | The rate for dropout layer | 0.2 |
| Training hyperparameters | | |
| LEARNING_RATE | The learning rate used in gradient descent formula | 0.001 |
| NUM_EPOCHS | The number of epochs to be used during model training | 300 |

# 3 SEQ2SEQ MODEL USING TRANSFORMER



# 3 SEQ2SEQ MODEL USING TRANSFORMER



# 3 COMPARISON BETWEEN ALL MODELS

| Model | MSE (on Test Set) |
|---|---|
| Seq2Seq LSTM (Auto-regressive) | *0.8637* |
| Seq2Seq LSTM (Non-Auto-regressive) | *0.8625* |
| Seq2Seq GRU (Auto-regressive) | *0.7817* |
| Seq2Seq GRU (Non-Auto-regressive) | *0.5784* |
| Transformer | *0.4640* |

**RANKING**

**1. TRANSFORMER (MSE = 0.4640)**
**2. SEQ2SEQ GRU (NON-AUTO-REGRESSIVE) (MSE = 0.5784)**
**3. SEQ2SEQ GRU (AUTO-REGRESSIVE) (MSE = 0.7817)**
**4. SEQ2SEQ LSTM (NON-AUTO-REGRESSIVE) (MSE = 0.8625)**
**5. SEQ2SEQ LSTM (AUTO-REGRESSIVE) (MSE = 0.8637)**

# THANK YOU!