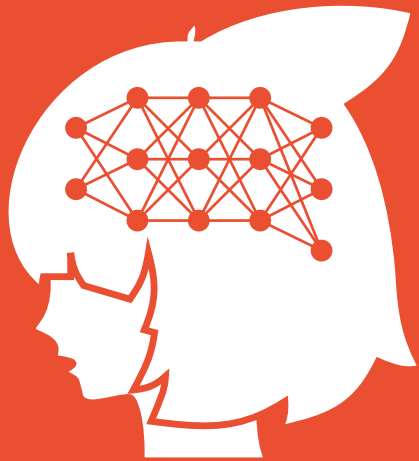


# 2024年VRChat自律機械知能プロジェクト 実験報告



ML Shukai

げそん<GesonAnko>

Myxy

Zassou

ぶんちん

田中スイセン

Earl Klutz (クルツ)

# 最初のお願い

- 本LTは視聴している皆さんを**Critic(直訳:映画の評論家)**と想定して作られています。
- **面白そうな点**や**よくわからない点**、**突っ込みどころ**など是非考えながら視聴してみてください！  
そして後で**ツッコミ**入れてください！

# 最初のお願い

- 本LTは視聴している皆さんを**Critic(直訳:映画の評論家)**と想定して作られています



うな点やよく

えなが

後でツッコ

難しい質問はげそんさん  
に丸投げするよ！  
やったね！

# VRChat自律機械知能プロジェクトとは

- VRChat上に**自律機械知能を生み出し**、その振る舞いを記録・解析することにより**人間及び知能全体への理解**を深める。
- また、それらを論文にまとめバ学会にて発表する。

# 去年の原始自律機械知能 P-AMI<Q> 実装

- Primitive Autonomous Machine Intelligence on Q(Cu)riosity.
- 原始的ではあるが、好奇心ベースの内発的動機付けを報酬に学習し探査し続けることが出来るP-AMI<Q>



# PAMI<Q>好奇心ベースの強化学習（予測誤差ベース）

## 1. 次に起こることを予測

Forward Dynamicsモデル  $f$

$f$  : 状態、行動  $\rightarrow$  次の状態

過去の経験から学習、予測誤差を最小化

## 2. 予測誤差（驚き）を報酬化

仮定：未学習なことは予測誤差が大きい

## 3. 報酬を最大化する行動を生成

Policy モデル  $\pi$

$\pi$  : 状態  $\rightarrow$  行動

強化学習の枠組みで学習



# PAMI<Q>好奇心ベースの強化学習（予測誤差ベース）

1. 次にを起

Forward

$f$ :状態

過去の経験

予想したものと実際のものが  
違えば違うほど高い報酬になる  
というのがポイントだよ！

2. 予 (き) を報酬化

ことは予測誤差が大きい

3. 化する行動を生成

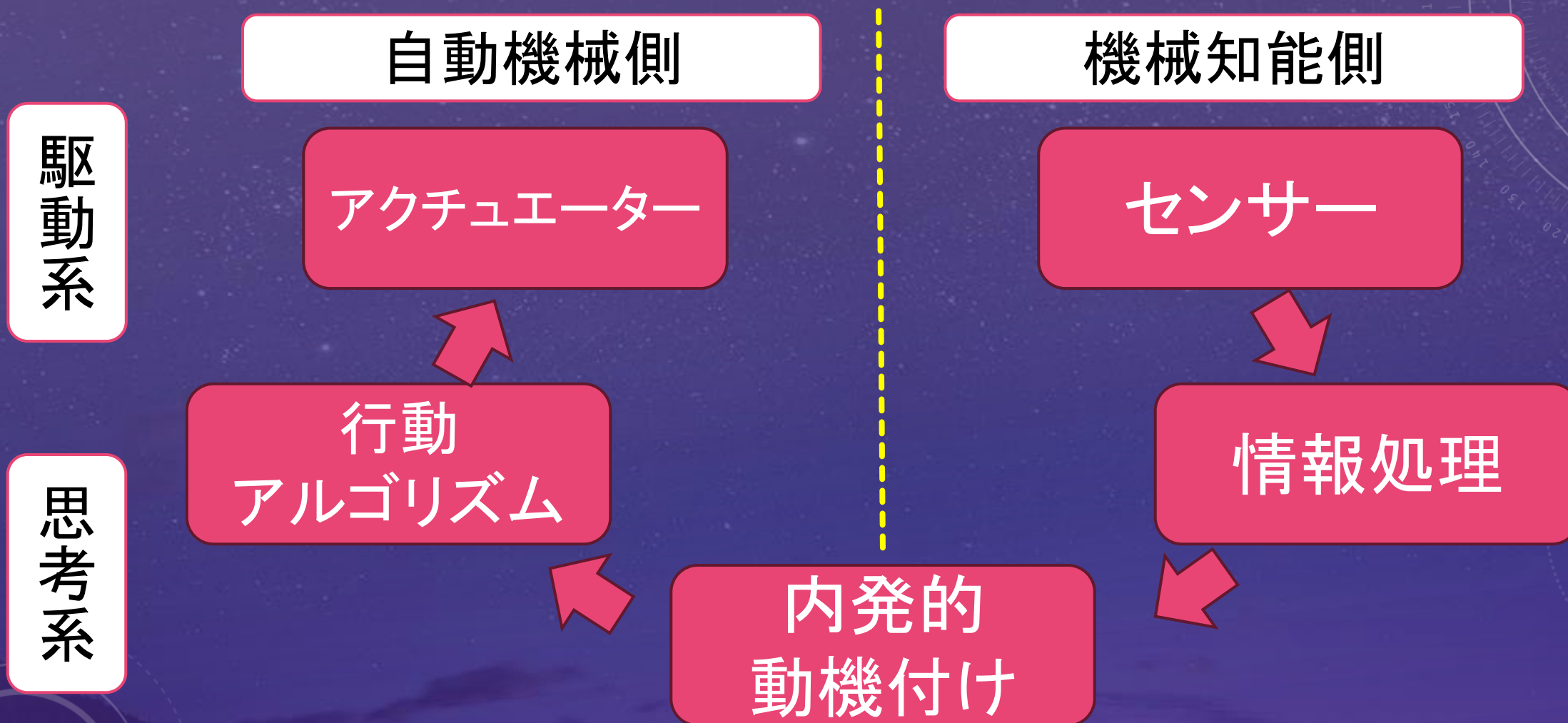
$\pi$

動

組みで学習



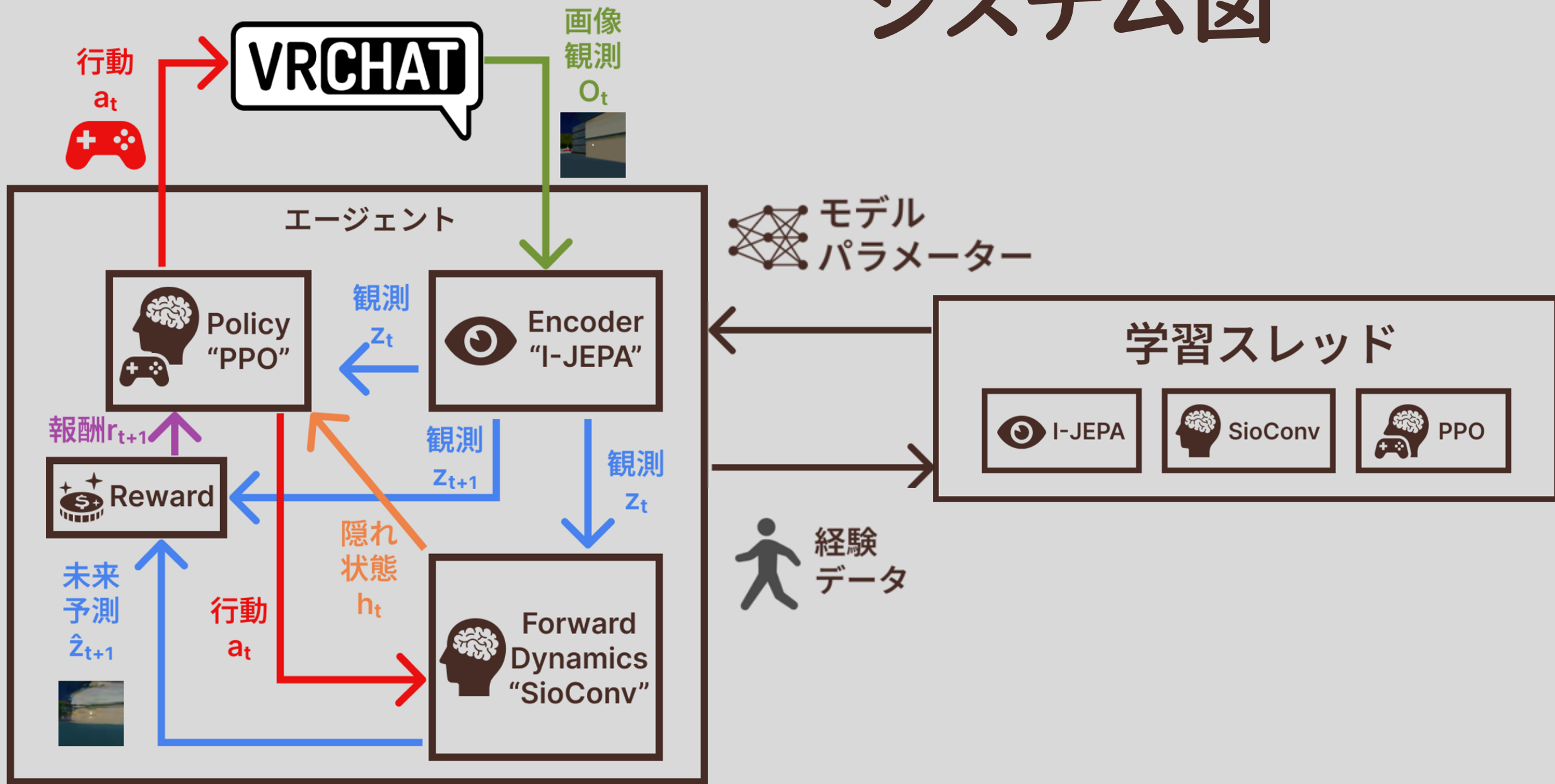
# VRChat上で動作する自律機械知能 基本構造





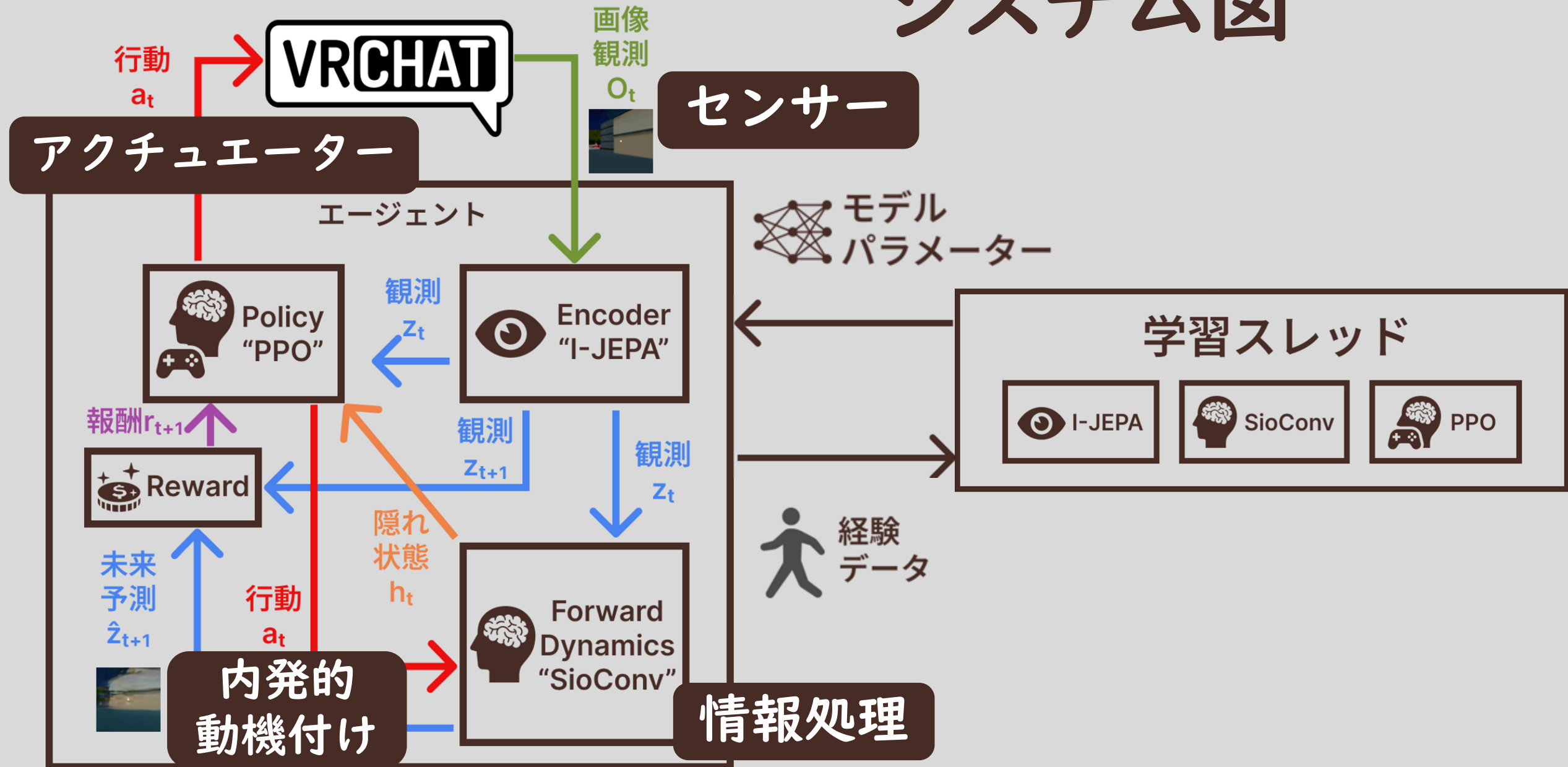
# 推論スレッド

# システム図



# 推論スレッド

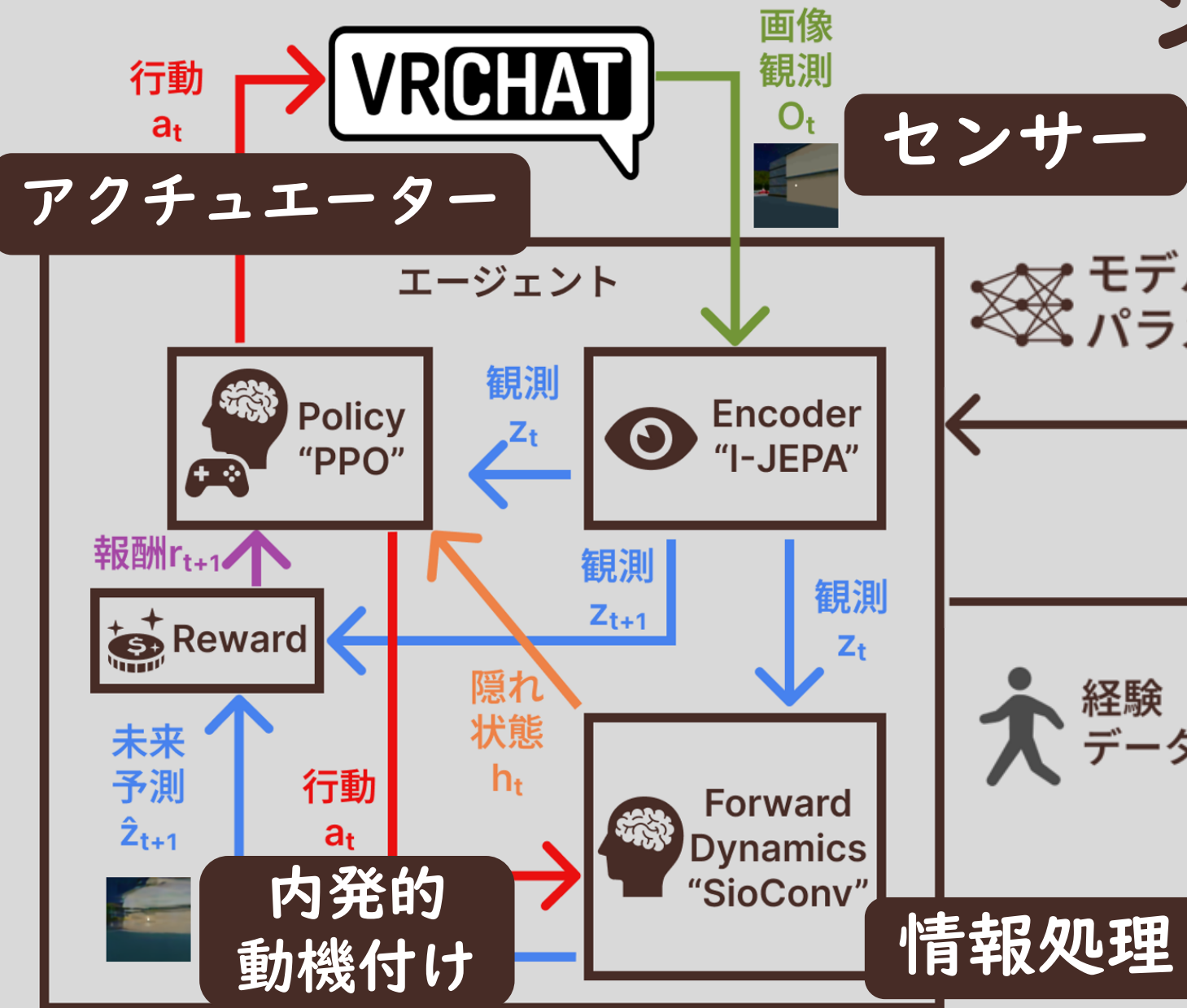
# システム図



# 推論スレッド

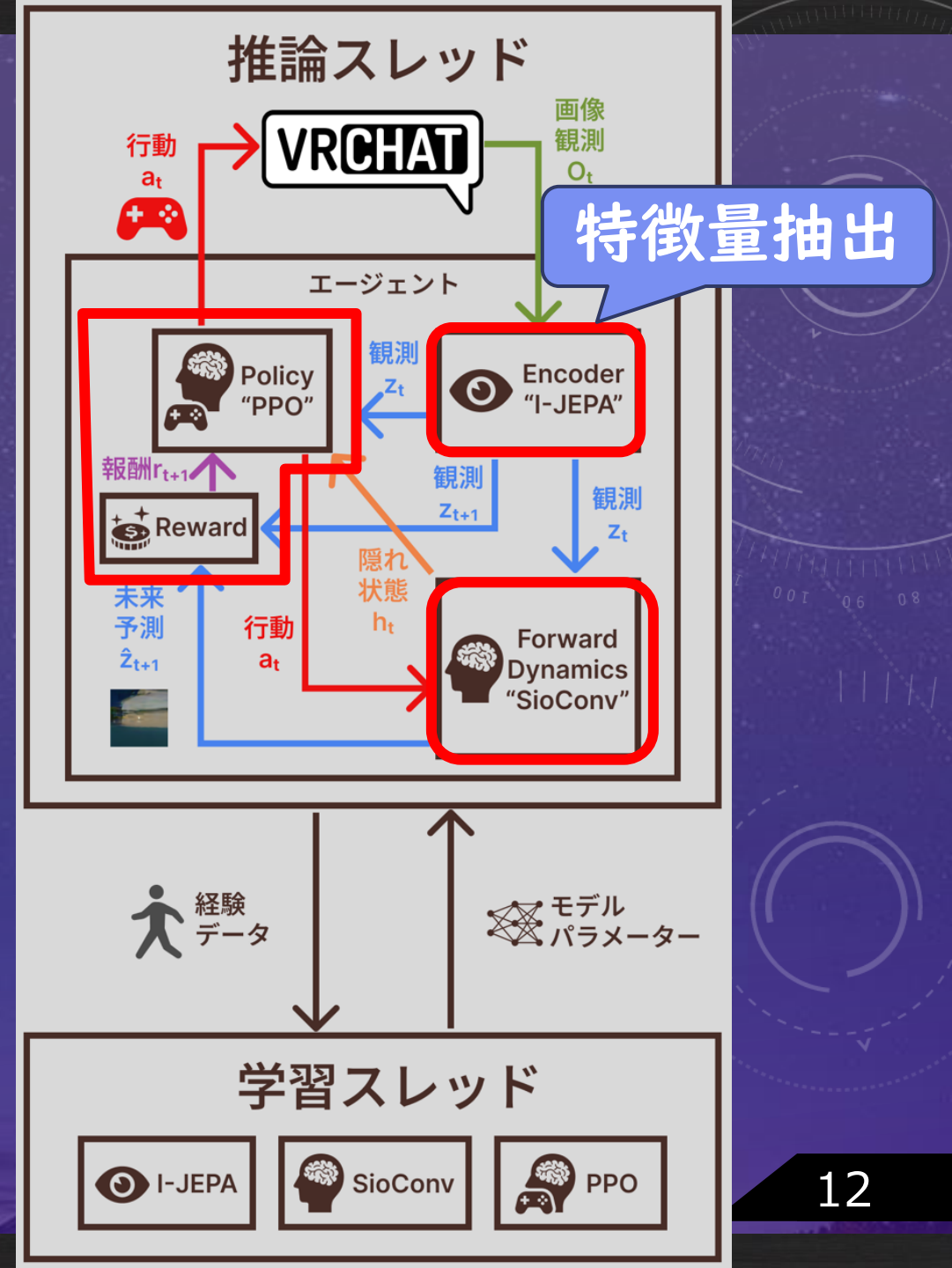
# システム図

必要なところだけ  
学習スレッドで学習してるよ!



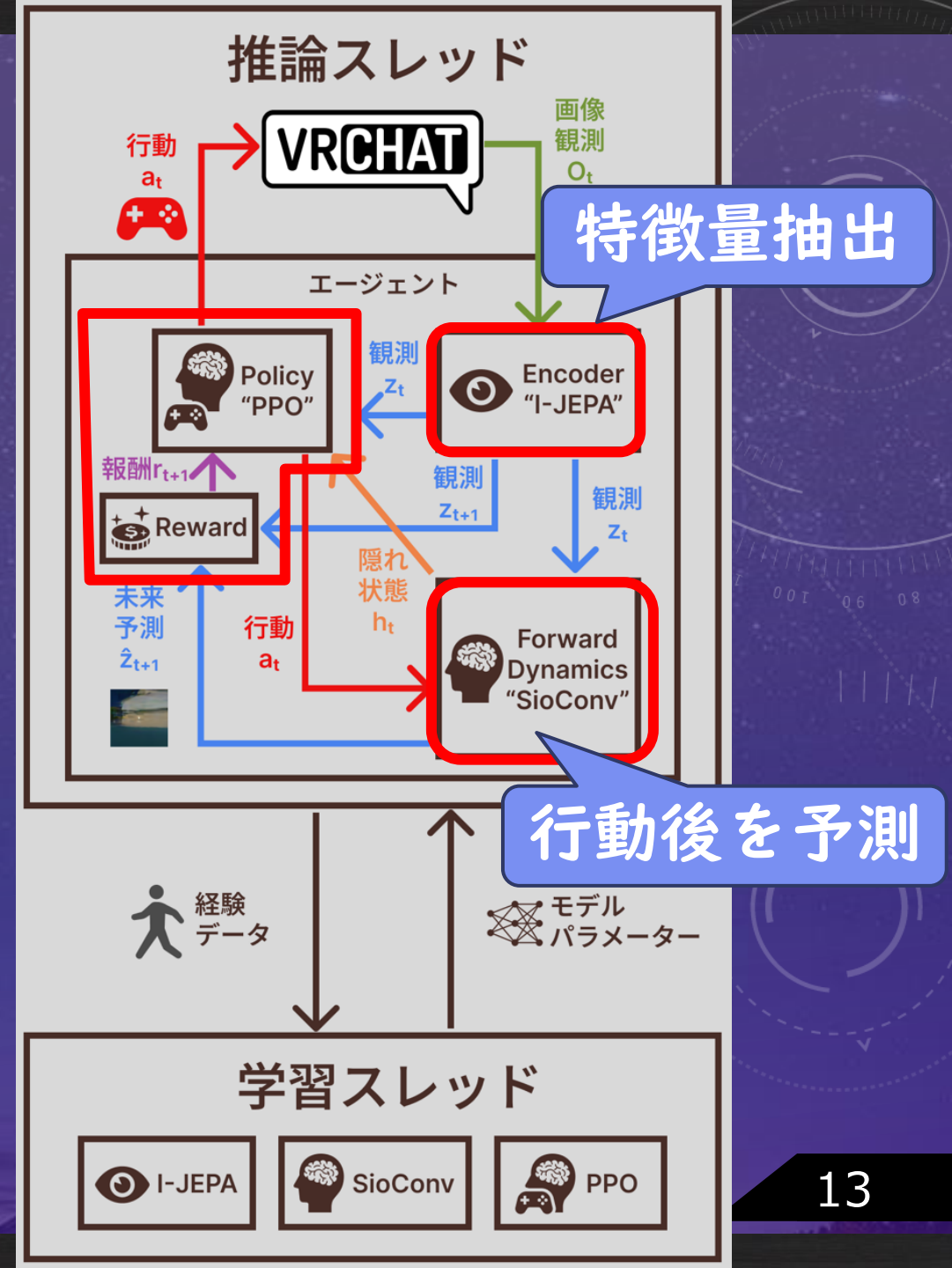
# P-AMI<Q>の基本構造

- 基本構造は内発的好奇心を報酬とした学習構造
- 観測Encoderで**特徴量抽出**



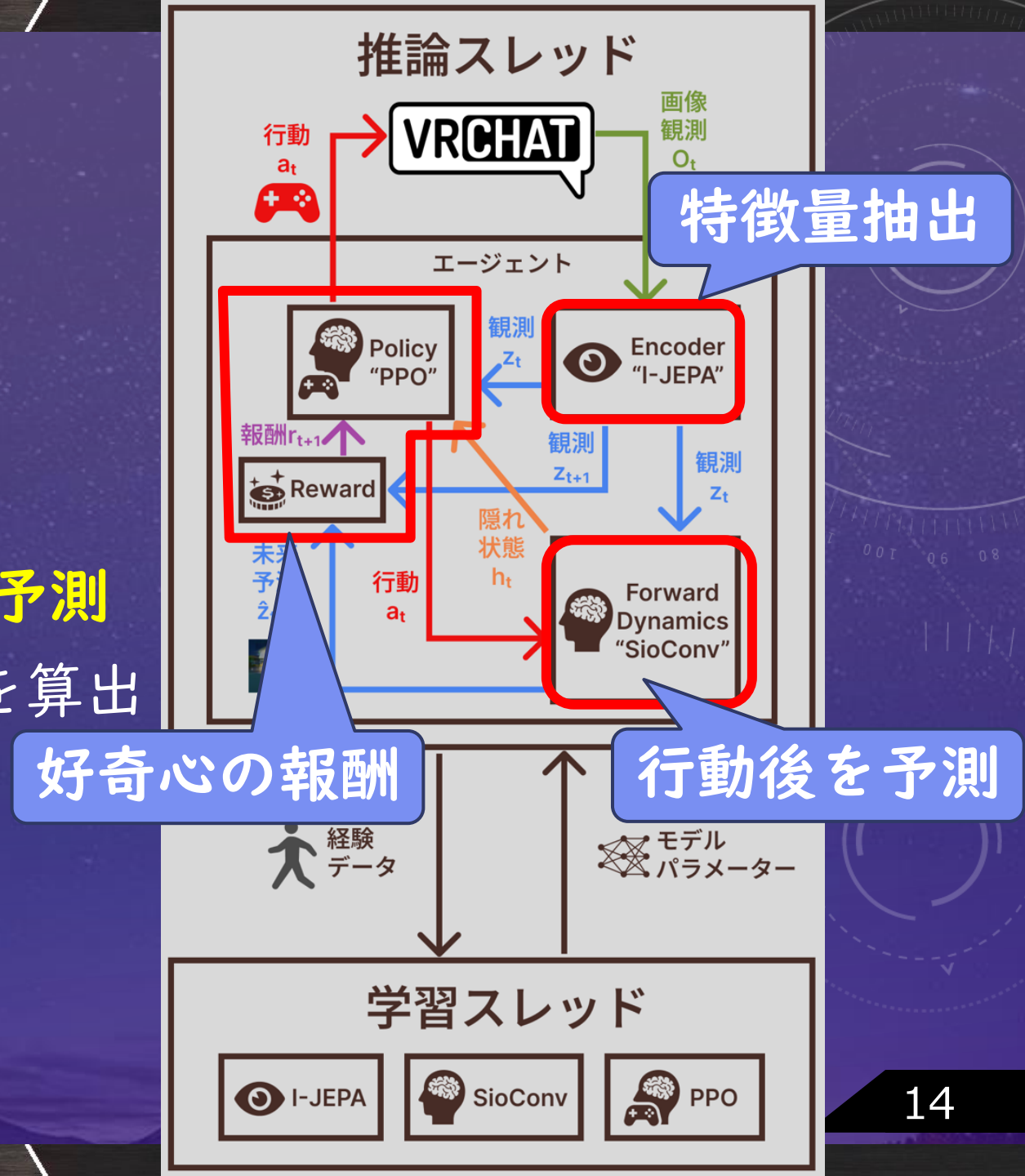
# P-AMI<Q>の基本構造

- 基本構造は内発的好奇心を報酬とした学習構造
  - 観測Encoderで**特徴量抽出**
  - Forward Dynamicsで**行動後を予測**



# P-AMI<Q>の基本構造

- 基本構造は内発的好奇心を報酬とした学習構造
  - 観測Encoderで**特徴量抽出**
  - Forward Dynamicsで**行動後を予測**
  - 特徴量と予測から**好奇心の報酬**を算出



# P-AMI<Q>の基本構造

- 基本構造は内発的好奇心を報酬とした学習構造
  - 観測Encoderで**特徴量抽出**
  - Forward Dynamicsで**行動後を予測**
  - 特徴量と予測から**好奇心の報酬**を算出
  - 強化学習で次の**行動を決定**



# P-AMT<0>の基本構造

内部のモデルは複雑になってきてるけど前から基本構造は同じだよ！

- 観測Encoderで**特徴量抽出**
- Reward Dynamicsで**行動後を予測**
- 予測から**好奇心の報酬**を算出
- 報酬で次の**行動を決定**

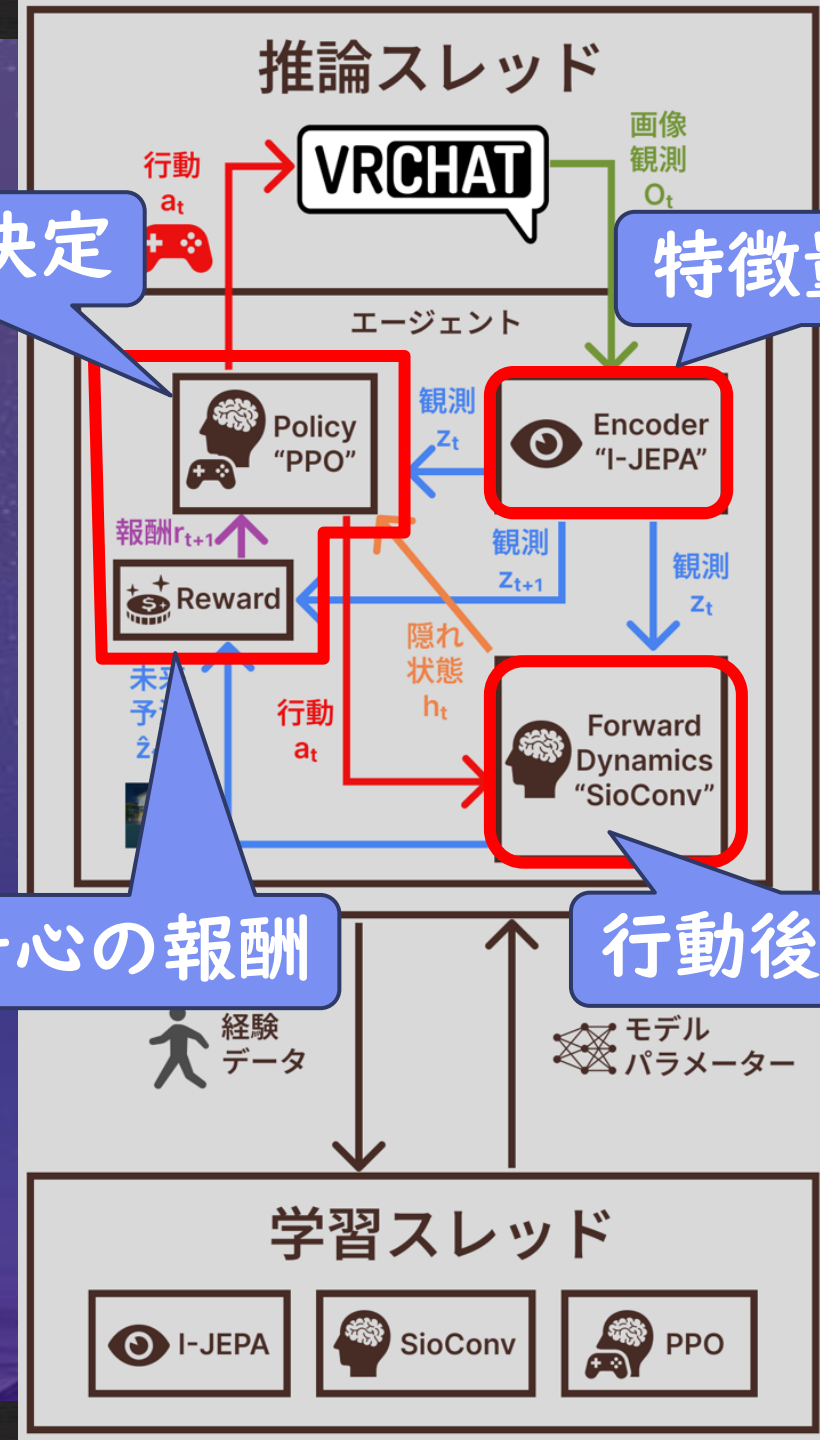


行動を決定

特徴量抽出

好奇心の報酬

行動後を予測





# I-JEPAとSioConvの学習と検証実験

- PAMI-<Q>自身の前にI-JEPAとSioConvを学習させ正しく動作するかを検証する必要がある
- 学習データの収集
  - Japan Street内の6地点でランダム行動するPAMI-<Q>を用いそれぞれ**4時間、計24時間分のデータ**を収集



# I-JEPAとSioConvの学習と検証実験

- PAMI-<Q>自身の前にI-JEPAと学習させ正しく動作するかを検証

学習データは  
集めるのは大変だよね！

- 学習データの収集

- Japan Street内の6地点でランダム行動するPAMI-<Q>を用いそれぞれ**4時間、計24時間分のデータ**を収集



# I-JEPAとSioConvの学習方法

- PAMI-<Q>と同じ非同期システムの形式で学習
- それぞれラージモデルとスモールモデルを用意
- Japan Streetで集めたデータからランダムにサンプリングして学習

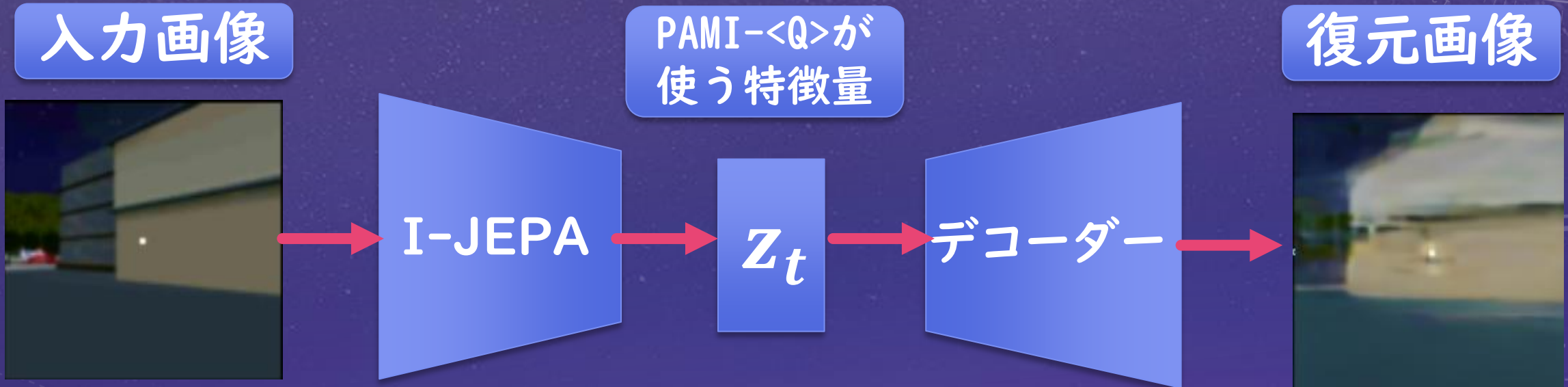


経験データを  
学習



# I-JEPAの検証方法

- エンコードされた画像を復元するモデル（デコーダー）を実装し、復元画像を目視確認。



# I-JEPA検証 生成画像と実観測のデータ

入力画像



復元画像  
(ラージモデル)



復元画像  
(スモールモデル)



# I-JEPA検証 生成画像と実観測のデータ

意外といい感じに  
復元できてるよ！

復元画像  
(ラベリング)

復元画像  
(スマートフォン)



# SioConvの検証方法

1. SioConvのみで学習可能かを確認
  2. 学習済みI-JEPAと同時学習
  3. 画像の時系列をわざとバラバラにした画像を使って学習（学習できないはず）
- 上記3フェーズで実施

# 各フェーズでのSioConvの検証

- 学習済みI-JEPAとデコーダ  
（同時学習時は未学習のもの）を用いる
- SioConvの生成した予測画像と観測画像を目視比較
- 学習中の損失値を計測



# SioConv検証 予測した生成画像と実観測のデータ

- ・ラージモデル

生成



実際

ステップ

1

6

11

16

21

26

# SioConv検証 予測した生成画像と実観測のデータ

- ・ スモールモデル

生成

実際



ステップ

1

6

11

16

21

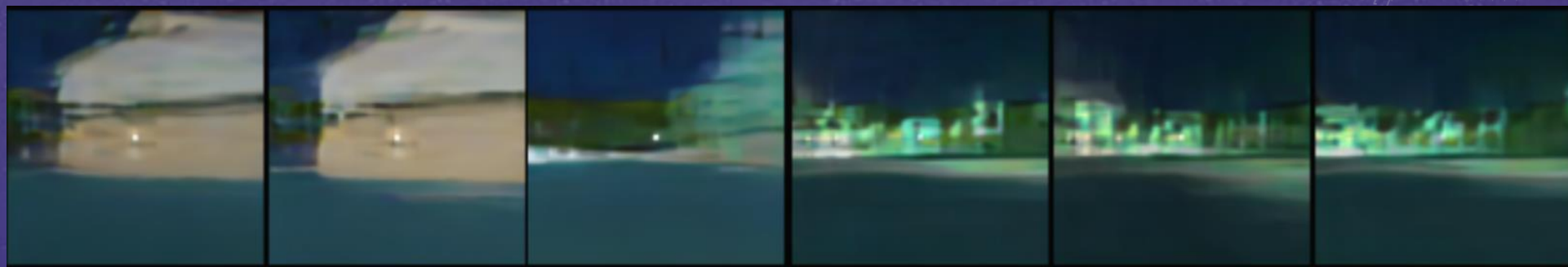
26

# SioConv検証 予測した生成画像と実観測のデータ

入力画像



復元画像  
(ラージモデル)

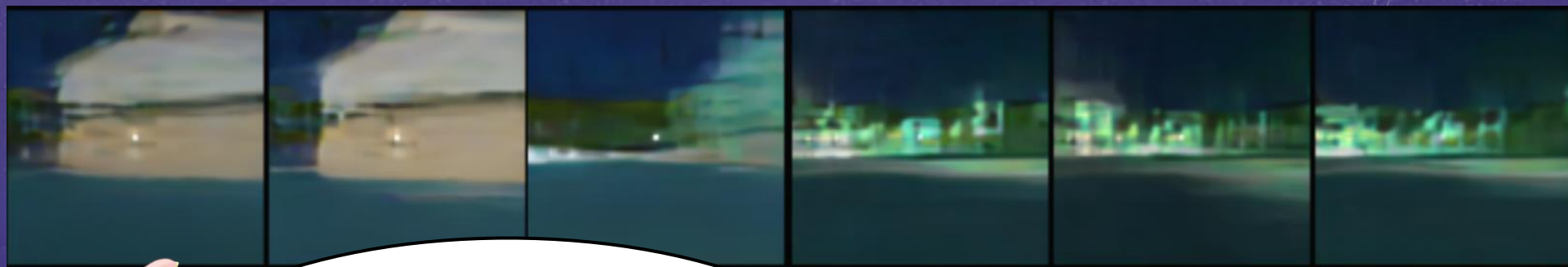


復元画像  
(スモールモデル)



# SioConv検証 予測した生成画像と実観測のデータ

入力画像



復元  
(=

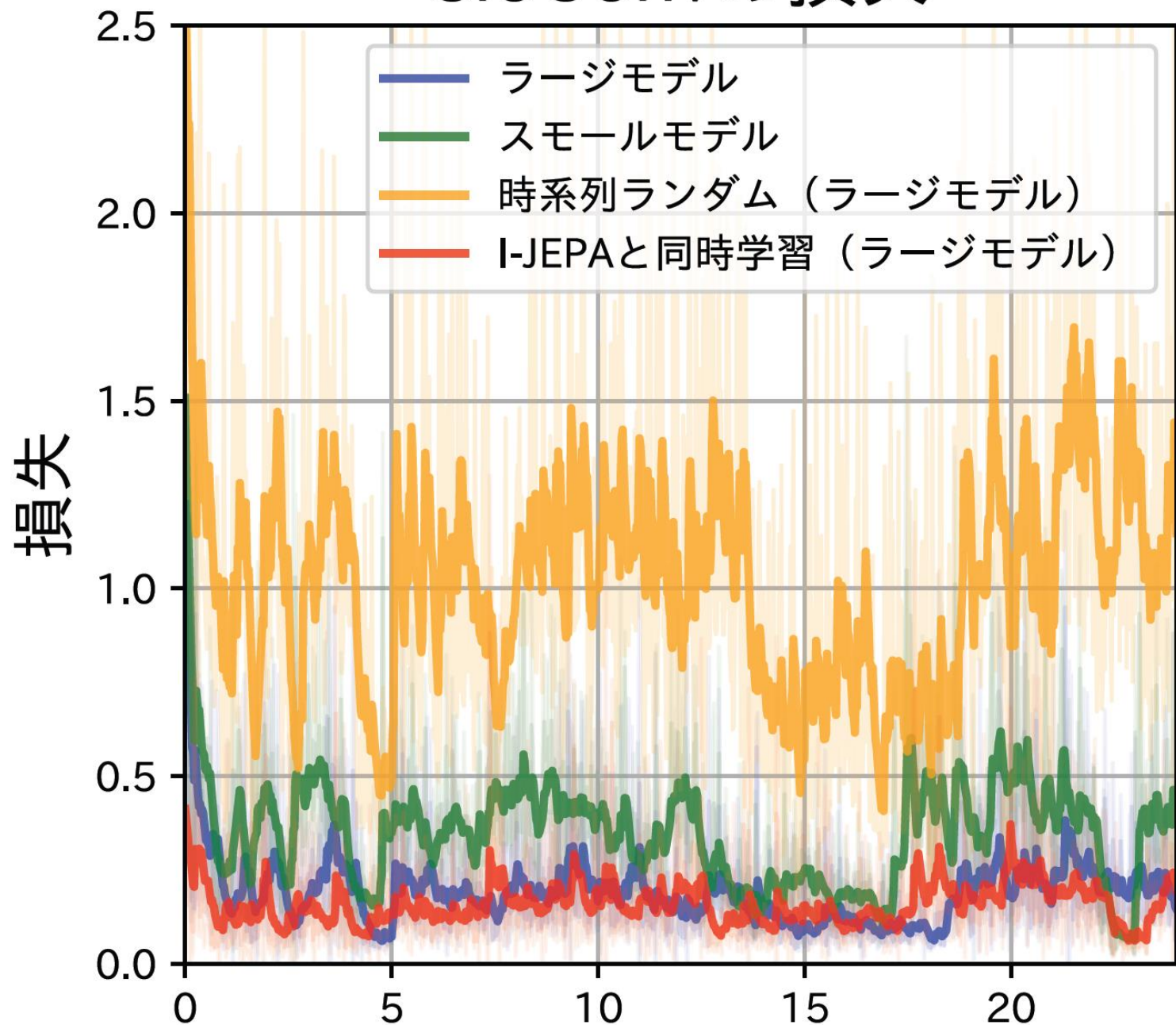
復元  
(=

予測画像も  
予想以上にはっきり  
見えててすごいよね！

# SioConvの損失

- ラージの方が損失は少ない
- 同時学習も問題なく学習できている
- ランダムは学習出来ていない

## SioConvの損失



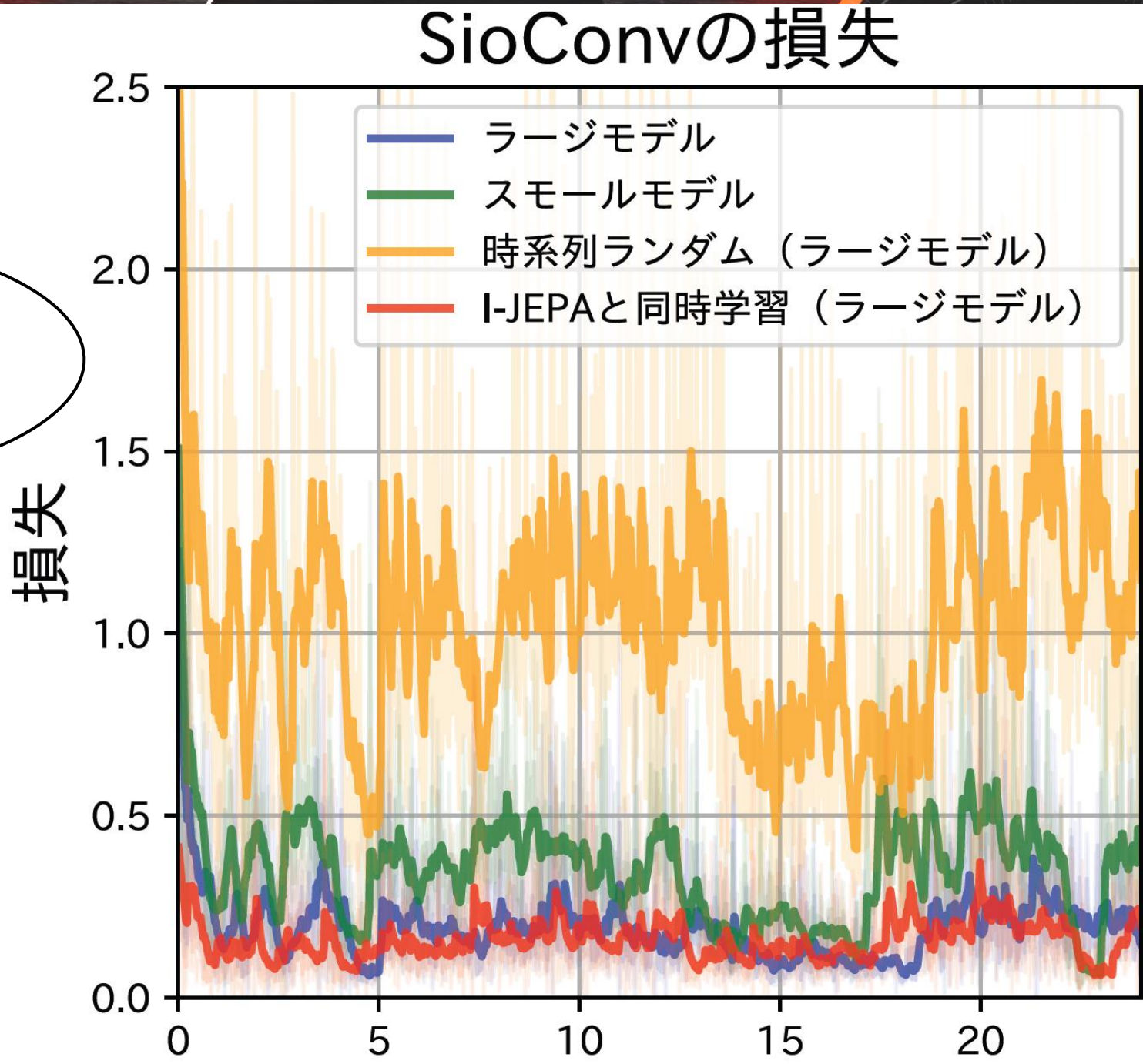
# SioConvの損失

- ラージの損失

損失もいい感じに  
仮説通りに動いて  
くれたよ！

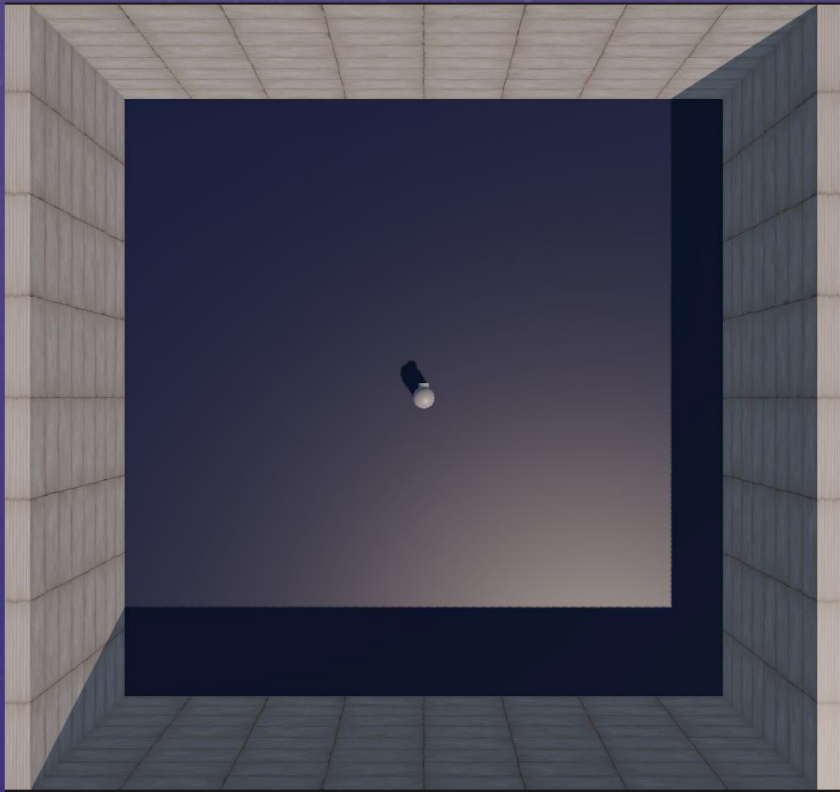
- 問題なく  
ている

- は学習  
な



# Policy(PAMI-<Q>)の検証

- 昨年と同様シンプルなワールドを使用

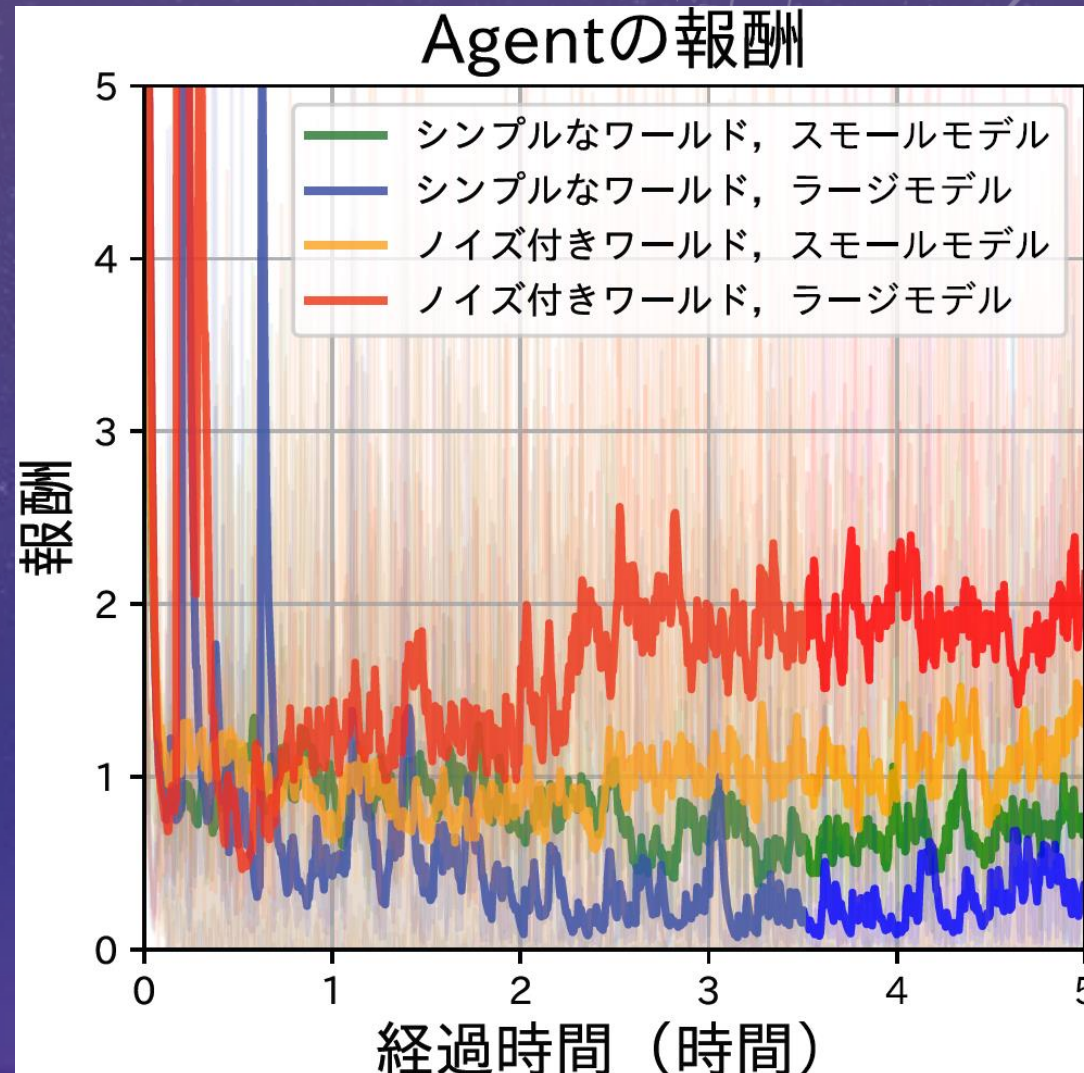
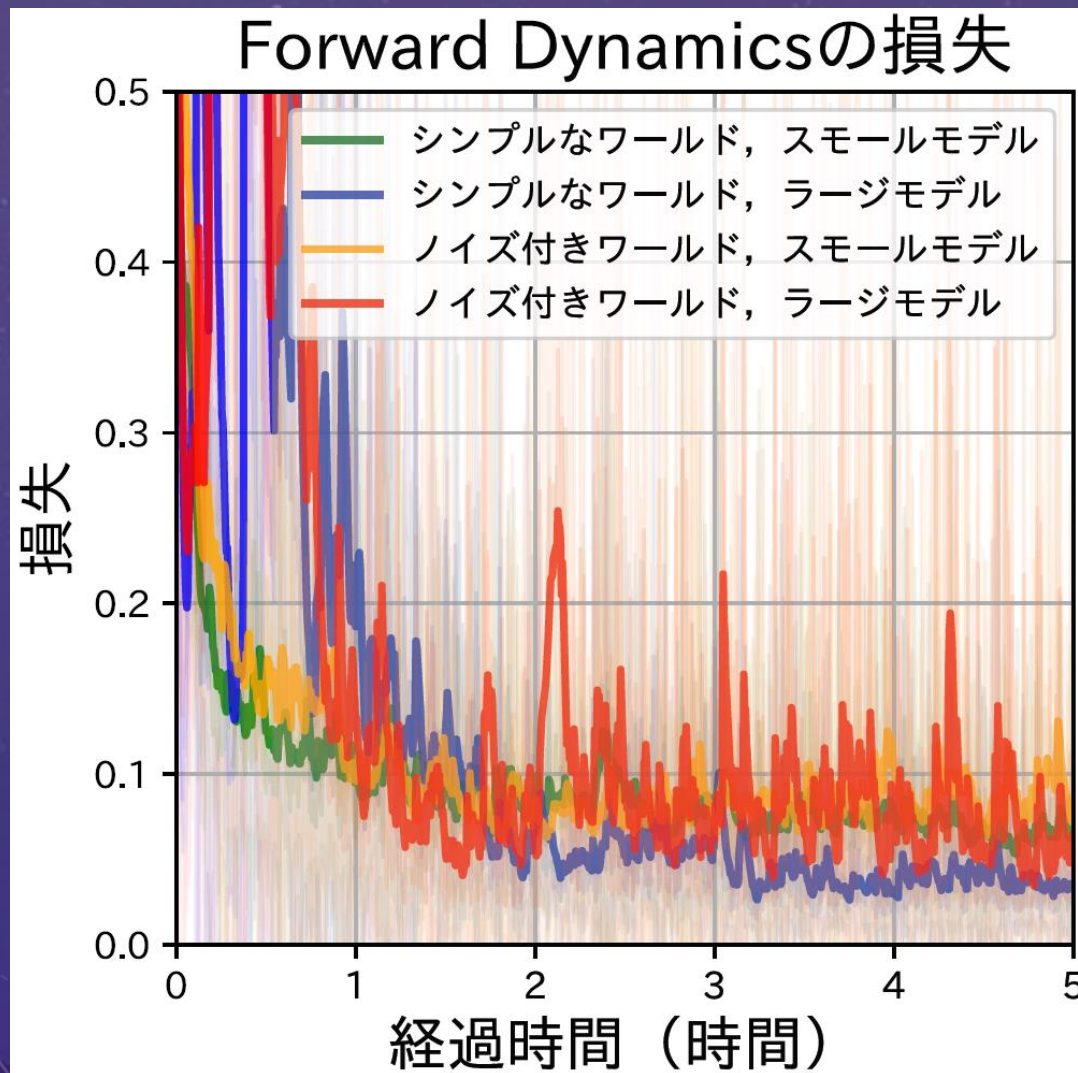


シンプルワールド



ノイズのあるシンプルワールド

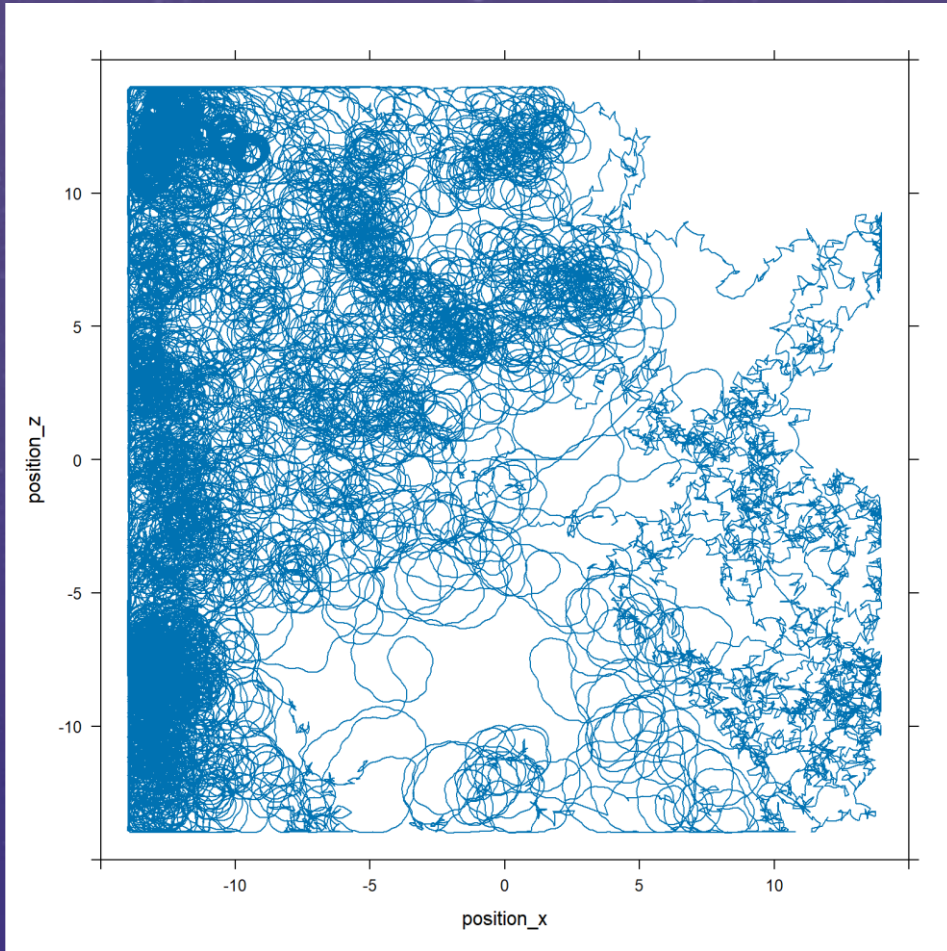
# PAMI-<Q>検証 各損失グラフ



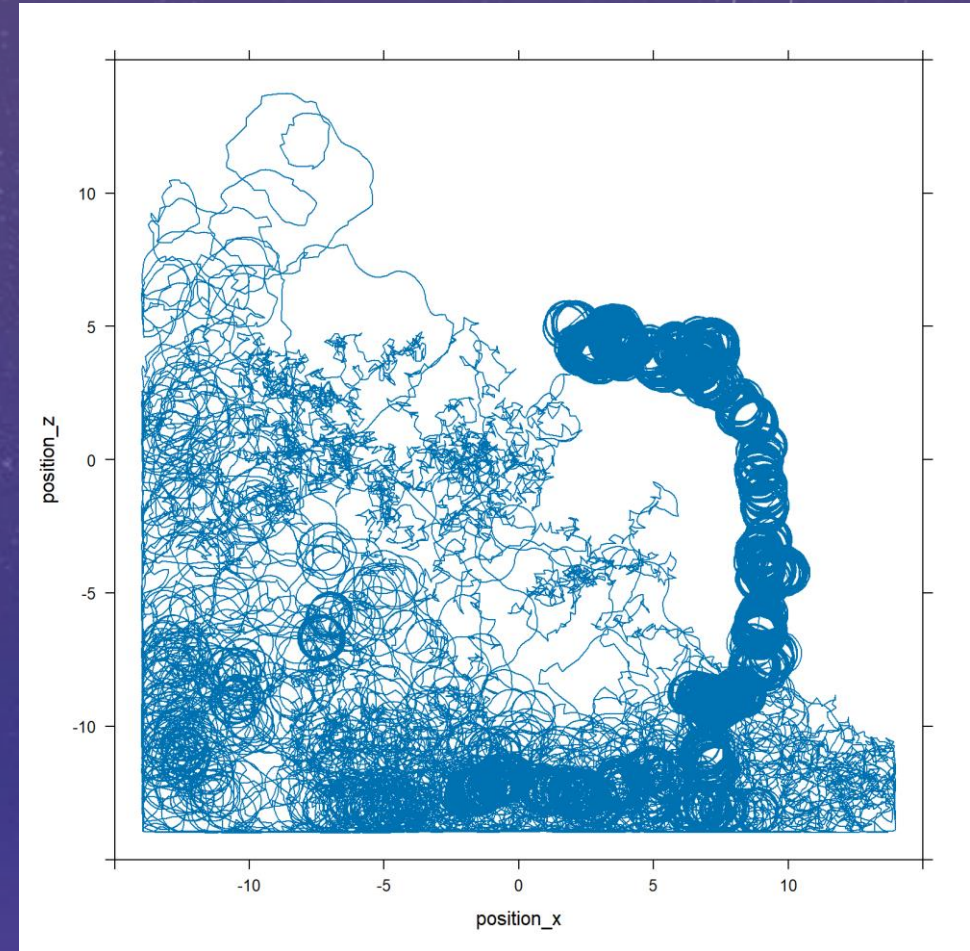
※ラージモデルで 5 回, スモールモデルで 10 回試行した結果  
の平均値を指数移動平均(平滑化係数:0.2)



# PAMI-<Q>検証 スモールモデルでの動き

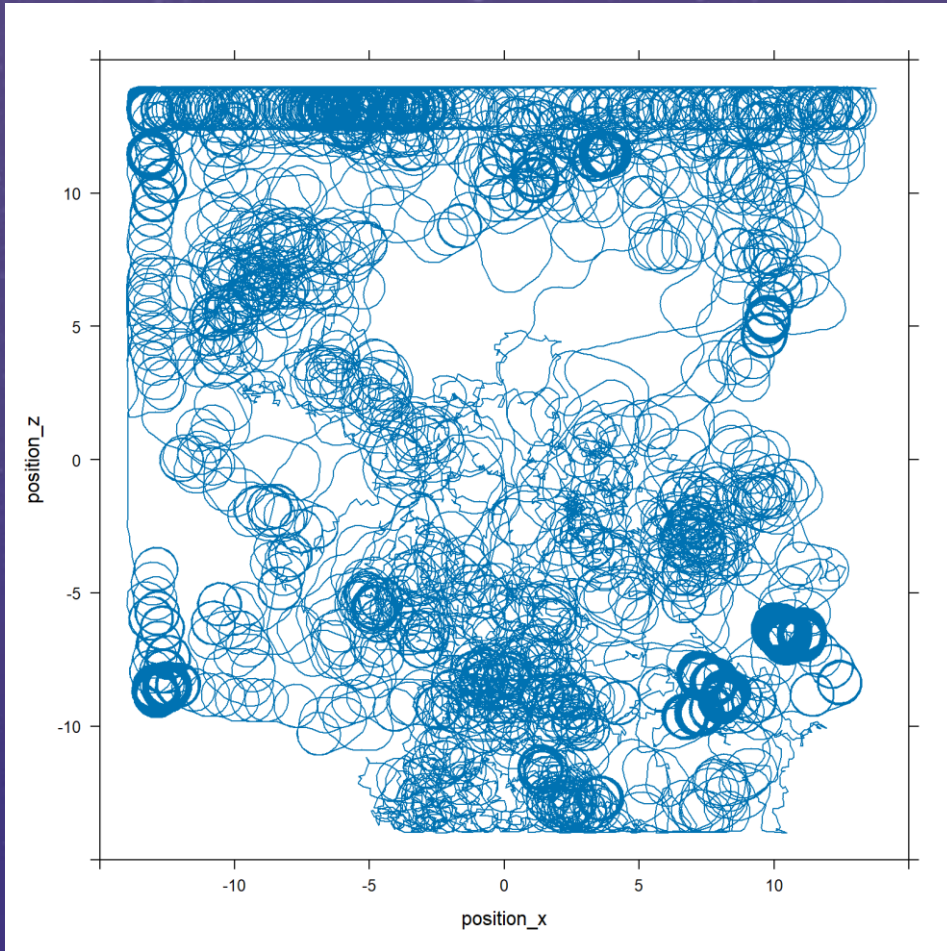


シンプルワールド

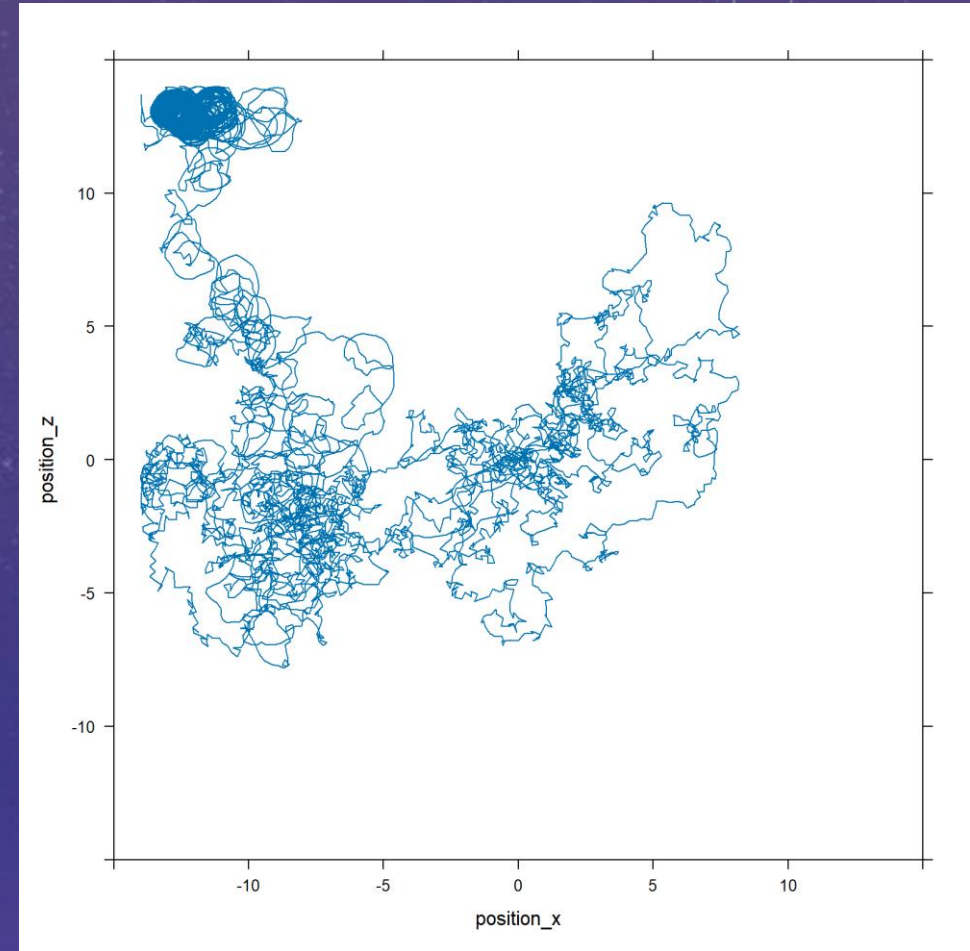


ノイズのあるシンプルワールド

# PAMI-<Q>検証 ラージモデルでの動き



シンプルワールド

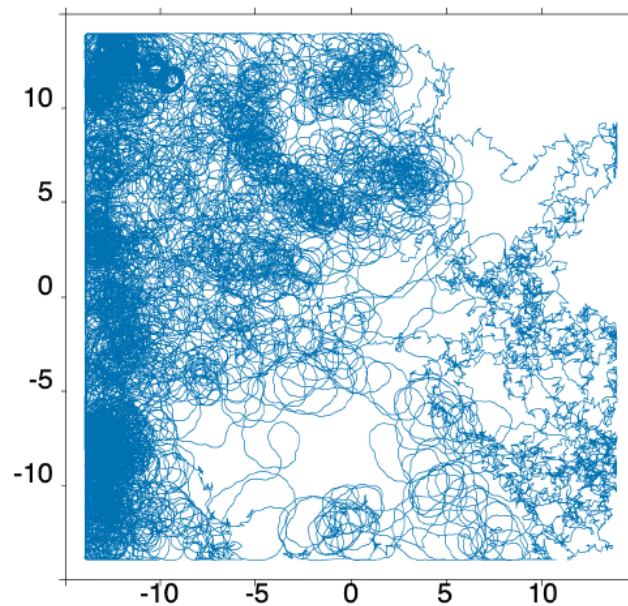


ノイズのあるシンプルワールド

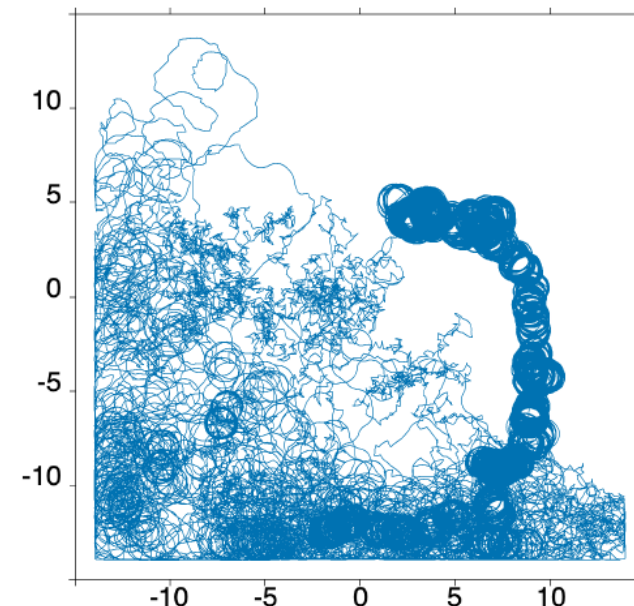
# PAMI- $\langle Q \rangle$ の動き

ノイズには  
影響されてるかな？

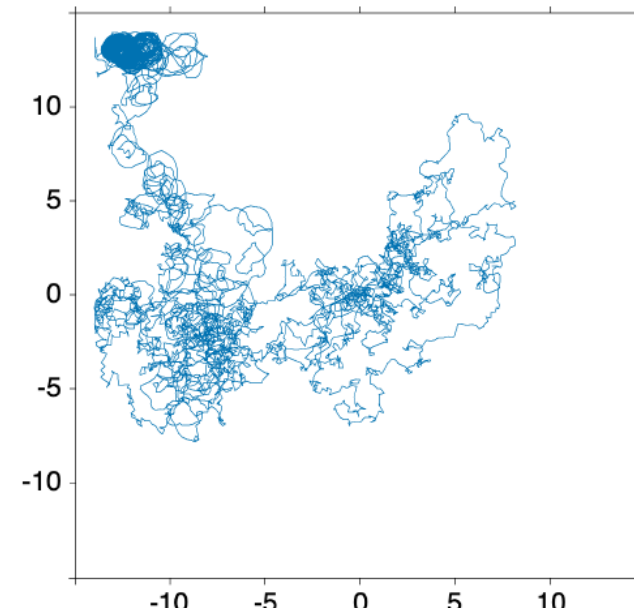
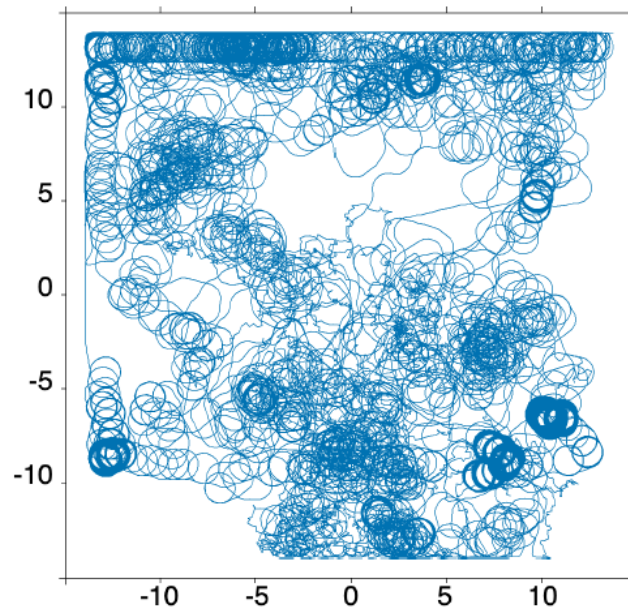
スモールモデル



ノイズ付きワールド



ラージモデル



# Policy (PAMI-<Q>) 検証まとめ

- ノイズ付きワールドの方が未到達が多いが報酬が得やすい為、ラージモデルで報酬は増大傾向

	シンプルなワールド	ノイズ付きワールド
Forward Dynamicsの損失	ラージ<スモール	ラージ<スモール
Agentの報酬	ラージ<スモール	ラージ>スモール

# Policy (PAMI-<Q>) 検証まとめ

- ノイズ付きワールドの方が未到達が多いが報酬が得やすい為、ラージモデルで報酬は増大傾向



Policyのモデルを  
スケールアップすると  
きちんと性能向上してる  
のがわかるよ

ラージ<スモール

ノイズ付きワールド

ラージ<スモール

ラージ>スモール

# 謝辞

- 機器・ネットワーク等提供  
ジーオーピー株式会社様



# 謝辞

- 機器・ネットワーク等提供  
ジーオーピー株式会社様



いつも  
ありがとうございます！



ML Shukai

ありがとうございます

By ML集会

げそん<GesonAnko>

Myxy

Zassou

ぶんちん

田中スイセン

Earl Klutz (クルツ)