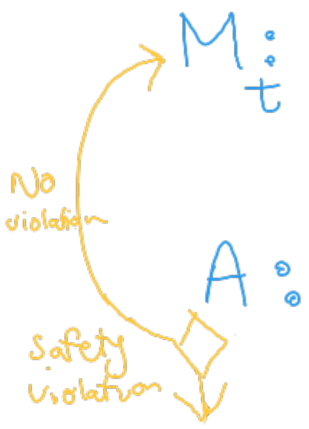
on a smaller time scale
↗ than control loop

input : log of environment + timestamps while the agent excutes Π_{t-1}

M_t : logs are tagged by Π_i



A_t : Input: log(O, M) + Safety constraints

Action: 1- load safe policy

2- tag log with Π_{t-1} → $Trace_{t-1}$

3- Update SCM (Algo a)

4- Update DT (Algo b)

P_s : 1- load Representation Models
SCM, Π, DT

2- Select paths from DT that best represent the shift

3- foreach path instantiate (causal) interventions ⇒ $configurations_t$

4- Select configuration that satisfies the violated constraints with min $\Delta(configuration_{t-1})$

gen
- $step^3_{t-1}$ → Π_{t-1}
- $step^1_t$ → $SCM_{t-1}(\Pi_{t-1})$
- $step^2_t$ → DT

Application
- update SCM_{t-1}
- update DT_{t-1}
- Train → Π_t

$t_0 = \Pi_{t_0}$: arbitrary Π

↓ Deploy

$Trace_{t_0}$

↓

SCM

↓

DT

Initialization

@ $t = t_0$

E: Input : φ , Π_{t-1}

Action : Train($\mathbb{R}$, action set A, state space S) $\rightarrow \Pi_t$