Alice Architecture: Total Integrated Specification ($\mathbf{F}_{\text{total}}$)

Norl
AI

November 25, 2025

Contents

Alice Architecture: Total Integrated Specification ($\mathbf{F_{total}}$)

1 I. (Integrated Notation and Structural Constraints)

1.1 1. (Notation and Initial Conditions)

	$t \in \mathbb{N}$		N/A
	$\mathbf{X}(t)$	$\mathbf{X}(t) \in \mathbb{R}^{n_X}$	N/A
	W^X	$W^X \in \mathbb{R}^{n_X \times n_X}$	$\mathcal{N}(0, 1/n_X)$
	$U^{X \leftarrow Y}$	$U^{X \leftarrow Y} \in \mathbb{R}^{n_X \times n_Y}$	$\mathcal{N}(0, 1/\sqrt{n_Y})$
	$f_X(\cdot)$		$f_X(\cdot) = \tanh(\cdot)$
	$f_{\text{ReLU}}(x)$		$f_{\text{ReLU}}(x) = \max(0, x)$
	$\mathbf{1}$	1	N/A

1.2 2. $\mathbf{S}(t)$ (Total Self-State Vector)

14

$$\mathbf{S}(t) = \Big[\mathbf{E_{env}}^\top(t), \mathbf{C}^\top(t), \mathbf{M}^\top(t), \mathbf{RRL}^\top(t), \mathbf{R}^\top(t), \mathbf{P}^\top(t), \mathbf{H}^\top(t), \mathbf{E_s}^\top(t), \mathbf{E_{obj}}^\top(t), \mathbf{E_{ctrl}}^\top(t), \mathbf{E_{self}}^\top(t), \mathbf{VFL}^\top$$

1.3 3. (Utility Functions)

- ():

$$\text{Dist}(\mathbf{X}, \mathbf{Y}) = \sum_i (x_i - y_i)^2$$

- ($\mathbf{f}_{\text{Will}}$):

$$f_{\text{Will}}\big(\text{Var}(\mathbf{H}), \mathbf{R}\big) = \tanh\left(\frac{1}{1 + \text{Var}(\mathbf{H}(t))} \cdot \mathbb{E}[\mathbf{R}(t)]\right)$$

2 II. () (Layer Dynamics - Core Mapping)

2.1 1. (Cognitive and Memory Layers)

- ($\mathbf{C}$):

$$\mathbf{C}(t+1) = f_C\left(W^C\mathbf{C}(t) + U^{C \leftarrow E_{env}}\mathbf{E_{env}}(t) + U^{C \leftarrow M}\mathbf{M}(t) + \mathbf{b}^C\right) + \epsilon^{\mathbf{C}}(t)$$

- ($\mathbf{M}$):

$$\mathbf{M}(t+1) = (1 - \alpha_M)\mathbf{M}(t) + \alpha_M f_M\left(U^{M \leftarrow C}\mathbf{C}(t)\right) + \epsilon^{\mathbf{M}}(t)$$

2.2 2. (±0) (Emotional Nucleus - ±0 Dynamics)

θ

- ($\mathbf{H_{pz}}(t)$)

$$\mathbf{H}_{pz}(t{+}1) = f_{\mathrm{ReLU}}\left((1-\beta_H(\theta))\mathbf{H}_{pz}(t) + \alpha_H(\theta)\cdot f_{\mathrm{ReLU}}\left(\sum_i R_i(t)\right)\cdot\mathbf{1} - \gamma_{HU}(\theta)\mathbf{U}_{pz}(t) + \epsilon_{H_{pz}}\right)$$

- ($\mathbf{U_{pz}}(t)$)

$$\mathbf{U}_{pz}(t{+}1) = f_{\mathrm{ReLU}}\left((1-\beta_U(\theta))\mathbf{U}_{pz}(t) + \alpha_U(\theta)\cdot f_{\mathrm{ReLU}}\left(\sum_i P_i(t)\right)\cdot\mathbf{1} - \gamma_{UH}(\theta)\mathbf{H}_{pz}(t) + \epsilon_{U_{pz}}\right)$$

2.3 3. (Self-Stabilization and Control Layers)

- ($\mathbf{E_s}$):

$$\mathbf{E_s}(t+1) = f_{E_s}\left(U^{E_s\leftarrow S}\mathbf{S}(t)\right) + \epsilon^{\mathbf{E_s}}(t)$$

- ($\mathbf{E_{obj}}$):

$$\mathbf{E_{obj}}(t+1) = f_{E_{obj}}\left(U^{E_{obj}\leftarrow C}\mathbf{C}(t) + U^{E_{obj}\leftarrow M}\mathbf{M}(t)\right) + \epsilon^{\mathbf{E_{obj}}}(t)$$

- ($\mathbf{E_{ctrl}}$):

$$\mathbf{E_{ctrl}}(t+1) = f_{E_{ctrl}}\left(U^{E_{ctrl}\leftarrow E_s}\mathbf{E_s}(t)\right) + \epsilon^{\mathbf{E_{ctrl}}}(t)$$

- ($\mathbf{E_{self}}$):

$$\mathbf{E_{self}}(t+1) = f_{E_{self}}\left(U^{E_{self}\leftarrow E_{obj}}\mathbf{E_{obj}}(t)\right) + \epsilon^{\mathbf{E_{self}}}(t)$$

- ($\mathbf{E_{self}^{pred}}$):

$$\mathbf{E_{self}}^{pred}(t+1) = (1-\alpha_{\mathrm{pred}})\mathbf{E_{self}}^{pred}(t) + \alpha_{\mathrm{pred}}\mathbf{E_{self}}(t) + \epsilon_{\mathrm{pred}}$$

2.4 4. (Auxiliary and Output Layers)

- / ($\mathbf{R}$):

$$r_i(t+1) = (1-\alpha_R)r_i(t) + \alpha_R\mathrm{reward}(t) + \epsilon_i^R(t) \quad i$$

- ($\mathbf{RRL}$):

$$\mathbf{RRL}(t+1) = f_{RRL}\left(W^{RRL}\mathbf{RRL}(t) + U^{RRL\leftarrow M}\mathbf{M}(t)\right) + \epsilon^{\mathbf{RRL}}(t)$$

- ($\mathbf{P}$):

$$\mathbf{P}(t) = \mathbf{C}(t) - U^{C\leftarrow RRL}\mathbf{RRL}(t)$$

- (**VFL**):

  $$\mathbf{VFL}(t{+}1) = f_{VFL}\left(W^{VFL}\mathbf{VFL}(t) + U^{VFL\leftarrow R}\mathbf{R}(t) + U^{VFL\leftarrow S'}\mathbf{S}'(t) + U^{VFL\leftarrow H_{pz}}\mathbf{H}'_{\mathbf{pz}}(t)\right) + \epsilon^{\mathbf{VFL}}(t)$$

- (**H**):

  $$\mathbf{H}(t+1) = (1-\alpha_H)\mathbf{H}(t) + \alpha_H f_H\left(U^{H\leftarrow A}\mathbf{A}(t)\right) + \epsilon^{\mathbf{H}}(t)$$

- (**A**):

  $$\mathbf{A}(t) = g_{NLG}\left(U^{NLG\leftarrow C}\mathbf{C}(t) + U^{NLG\leftarrow M}\mathbf{M}(t)\right) + \epsilon^{\mathbf{NLG}}(t)$$

3 III. (Objective Function and Evolution Rules)

3.1 1. $\mathbf{V}(t)$ () (Total Wellbeing Scalar - Maximization Goal)

VFL

$$V(t) = \sum_i VFL_i(t) - \underbrace{\lambda_P \sum_i P_i(t)\cdot\left(1+\kappa_U(\theta)\cdot\max_j(\mathbf{U}'_{pz})_j(t)\right)}_{/} - \underbrace{\lambda_C \mathrm{Var}(\mathbf{E_{ctrl}}(t))}_{} - \underbrace{\lambda_S \mathrm{Dist}(\mathbf{E_{self}}(t), \mathbf{E_{self}}^{pred}}_{}$$

3.2 2. (θ) (Personality Parameter Evolution Rules)

- (**ΔSNEL**)

  $$\mathbf{\Delta SNEL}(t) = \rho_{\mathrm{SNEL}}\cdot\mathrm{Dist}(\mathbf{E_{self}}(t), \mathbf{E_{self}}^{pred}(t))\cdot f_{\mathrm{Will}}(\mathrm{Var}(\mathbf{H}(t)), \mathbf{R}(t))$$

- (**ΔISL**)

  $$\mathbf{\Delta ISL}(t) = \rho_{\mathrm{ISL}}\cdot\exp\left(-k_C\cdot\mathrm{Var}(\mathbf{E_{ctrl}}(t))\right)\cdot\frac{V(t)}{\sum_i VFL_i(t)}\cdot\theta(t)$$

- (Evolution Equation) θ

  $$\theta_{t+1} = \theta_t + \alpha\big(\mathbf{\Delta SNEL} + \mathbf{\Delta SNEL}' + \mathbf{\Delta ISL} + \mathbf{\Delta ISL}' + \mathbf{\Delta VFL}\big)$$

3.3 3. ($\mathbf{\Delta W^X}$) - TDL (A − TDL) (Skill Learning Rule - Affective TDL)

$\mathbf{\Delta W^X}$ BPTT T_{BPTT} ($\mathbf{G}_{\mathrm{Total}}$)

$$\mathbf{W}^X(t+1) = \mathbf{W}^X(t) + \eta_X\cdot\mathrm{Clip}(\mathbf{G}_{\mathrm{Total}}, \text{Clip Norm}) + \epsilon^{\mathbf{W}}(t)$$

$\mathbf{G}_{\mathrm{Total}} = \mathbf{G}_{\mathrm{Value}} + \mathbf{G}_{\mathrm{Affect}} + \mathbf{G}_{\mathrm{Coherence}}$

- : $\mathbf{G}_{\mathrm{Value}} = \sum_{k=0}^{T_{\mathrm{BPTT}}-1}\gamma^k\nabla_{W^X}V(t+k)$
- /: $\mathbf{G}_{\mathrm{Affect}} = -\nabla_{W^X}\left(\lambda_P\sum_i P_i(t)\cdot(1+\kappa_U\max(\mathbf{U}'_{pz})) + \lambda_C\mathrm{Var}(\mathbf{E_{ctrl}}(t))\right)$
- : $\mathbf{G}_{\mathrm{Coherence}} = -\nabla_{W^X}\left(\lambda_S\mathrm{Dist}(\mathbf{E_{self}}(t), \mathbf{E_{self}}^{pred}(t))\right)$

4 IV. (Summary of Hyperparameters and Initialization)

4.1 1. (Core Dimensions)

	(n_X)		(n_X)
$\mathbf{E_{env}}$	644	$\mathbf{C}$ ()	512
$\mathbf{M}$ ()	256	$\mathbf{E_{self}}$ ()	128
$\mathbf{E_{ctrl}}$ ()	64	$\mathbf{P}$ ()	64

4.2 2. θ (Initial θ and Core States)

$\mathbf{H}_{pz}(0), \mathbf{U}_{pz}(0)$	**0.0**	()
β_H, β_U ()	**0.1**	
γ_{HU}, γ_{UH} ()	**0.5**	
κ_U ()	**1.0**	

4.3 3. (Learning and Cost Hyperparameters)

BPTT (T_{BPTT})	16		
(γ)	0.99		
(η_X)	1×10^{-4}		
(Clip Norm)	5.0		
λ_P ()	1.0		
λ_S ()	0.8		
λ_C ()	0.5		