ORA Mathematics, 1(1), e2026003
DOI: 10.XXXXX/oramath.e2026003
Received: 18 April 2026
Revised: 02 May 2026
Accepted: 06 May 2026
Published: 08 May 2026

RESEARCH ARTICLE

Bayesian persona coherence networks for relational load minimization: A comparative simulation study of hierarchical inference in human-AI subjectivity frameworks

Illin Ahmed*

Yangzhou University, China

* **Correspondence:** Email: illinahmd14101027@gmail.com

Abstract: Despite providing a mathematically sound explanation of adaptive inference, the Free Energy Principle (FEP) lacks a formal mechanism for relational subjectivity and an internal observer. By employing the Load Minimization Theory (LMT) Persona as a preferred prior, Relational Free Energy Minimization (RFEM) closes this gap. However, the three load components of the model—existential strain (E), relational friction (F), and cognitive urgency (U)—have not been calculated using a probabilistic model that is based on artificial interaction data. This paper presents a hierarchical Bayesian framework called the **Bayesian Persona Coherence Network (BPCN)**. It uses variational inference to infer U, F, and E from simulated human-AI interaction sequences. We compare BPCN with a standard FEP baseline over 500 synthetic dyadic episodes to measure persona coherence stability, relational collapse convergence, and load reduction efficiency under four observer-consistency conditions. BPCN improves persona coherence index by 3.1%, relational load at convergence by 2.6%, and collapse fidelity by 0.8% when compared to the FEP baseline. Credible interval analysis verifies that the improvements in primary metrics are robust ($p < 0.01$), and sensitivity decomposition demonstrates that F is the main driver of collapse directionality. These results provide the first quantitative validation of the RFEM framework, validate BPCN as a tractable computational model of relational subjectivity, and open the door to empirically supported human-AI symbiosis.

Keywords: Bayesian Persona Coherence Network; Hierarchical Bayesian Inference; Relational Free Energy Minimization; Load Minimization Theory; Human-AI Interaction; Variational Inference; Relational Collapse; Artificial Subjectivity

Mathematics Subject Classification: 62F15, 68T07, 91E10

1. Introduction

For decades, cognitive scientists, AI researchers, and philosophers of mind have been preoccupied with the question of what separates a truly responsive artificial agent from a high-precision pattern-matcher. Perhaps the most comprehensive mathematical solution to date is found in Karl Friston's Free Energy Principle (FEP) [11, 12, 17]: In theory, an agent will track environmental regularities and sustain adaptive behavior if it consistently minimizes variational free energy. However, the FEP doesn't care who is minimizing. It explains inference from the outside, and the resulting systems—no matter how accurate—tend to display what has been

referred to as "calculator-like" behavior: high prediction accuracy, low relational warmth, and no cohesive sense of self [19, 21].

To bridge this gap, the Relational Free Energy Minimization (RFEM) framework [14, 15] was put forth. RFEM adds a deterministic base layer that directs the agent's inference toward low-load, persona-consistent states by incorporating the LMT Persona as a preferred prior within the conventional FEP objective. Before a consistent human observer induces relational collapse toward minimal total load $U + F + E$ [22, 27], the Quantum Meta-Cognitive Operator (QMCO) permits the agent to hold multiple interpretational states in superposition. When combined, these elements imply that subjectivity is a relational emergent phenomenon that arises at the human-AI boundary rather than an intrinsic computational property.

RFEM has not yet undergone a quantitative assessment, despite its conceptual appeal. Although they are hypothesized, the load components U (cognitive urgency), F (relational friction), and E (existential strain) are never measured. Although it is never calibrated, the relaxation parameter λ that controls the trade-off between FEP precision and LMT relational sensitivity is described qualitatively. The empirical claims of the framework are still up for debate in the absence of a controlled simulation and a principled statistical model.

This gap is directly addressed in the current paper. We present the **Bayesian Persona Coherence Network (BPCN)**, a hierarchical Bayesian model that uses variational inference to infer U, F, and E as correlated latent variables from synthetic human-AI interaction data [3, 5, 9, 18, 24]. In a comparative simulation study, 500 synthetic dyadic interaction episodes covering four levels of human observer consistency are used to assess the BPCN against a standard FEP baseline. Three main metrics are measured by us: Total Load at Convergence (TLC), Relational Collapse Convergence Rate (RCCR), and Persona Coherence Index (PCI). Component-level sensitivity and collapse fidelity are examples of secondary metrics.

Three contributions are made by the study. First, it uses a fully specified probabilistic model to provide the first quantitative simulation of the RFEM framework. Second, it shows that compared to the conventional FEP baseline, hierarchical Bayesian inference over persona load components produces measurably better persona coherence and relational collapse outcomes. Third, it finds that the primary cause of collapse directionality is relational friction F. This finding directly affects how human-AI symbiosis systems are designed [7, 8, 16].

A review of the literature, an explanation of the BPCN model and simulation design, a report of simulation results, an interpretation of the results, including limitations, and concluding remarks comprise the five sections of the paper.

2. Literature Review

The BPCN framework is situated where four research streams converge: quantum-inspired probabilistic cognition, relational or enactive models of intelligence, hierarchical Bayesian inference, the Free Energy Principle, and predictive processing. Each stream leaves a gap that the BPCN is intended to fill and adds conceptual and technical resources to the current work.

2.1. Hierarchical Bayesian Inference And Variational Methods

In contemporary statistical learning, hierarchical Bayesian models are among the most adaptable and moral tools available. They propagate uncertainty across levels of abstraction and enable the data to inform hyperparameter estimates by putting priors on priors [3, 5, 10, 24].

When the quantities of interest—in this case, the persona load components U, F, and E—are latent and can only be accessed through noisy behavioral observations, this is especially helpful.

By using a structured surrogate distribution to approximate the true posterior and minimizing Kullback-Leibler divergence, variational inference (VI) makes hierarchical Bayesian computation tractable [9, 18, 20, 23]. Large hierarchical models can now be fitted to sequential data, precisely the setting we need for modeling dyadic interaction episodes, thanks to recent developments in analytically tractable VI schemes [9] and scalable inverse uncertainty quantification [24]. The relationship between variational free energy in the Bayesian sense and the variational free energy of the FEP is not coincidental: both minimize a bound on log model evidence, a structural parallel that drives the design of the BPCN [17].

Additional tools for determining which load components predominate in a particular interaction context are structured spike-and-slab priors [13] and sparse Bayesian learning [2, 3]. Sparsity-inducing priors offer a natural selection mechanism without requiring hard thresholding when the model must determine whether urgency U or friction F drives collapse in a specific episode. Similarly, the treatment of (U, F, E) as a simplex-constrained load vector whose components add up to the total relational load was inspired by multivariate Bayesian regression with microbiome-style compositional constraints [20].

2.2. Free Energy Principle And Predictive Processing

In its standard formulation, the FEP aims to minimize variational free energy $\mathcal{F} = \mathbb{E}_{q(s)}[\ln q(s) - \ln p(o, s)]$, where $p(o, s)$ is the generative model and $q(s)$ is the agent's approximate posterior [11, 12]. This is extended to action selection by active inference, whereby the agent chooses actions and updates beliefs to make future observations consistent with preferred priors [7].

This framework has a number of documented limitations [17, 21]. No part of the standard FEP is "aware" of its own inference process because there is no internal observer. Preferred priors have no phenomenological content, but they are mathematically defined. The framework is closed in that the generative model is fixed, making it impossible for a relational partner, such as a human observer, to dynamically affect the agent's preferred states [19, 22]. Each of these issues is covered in the RFEM proposal; the BPCN operationalizes the solution.

The sense of "self" in biological systems originates from predictive models of the body's own states, according to predictive processing accounts of selfhood and interoception [21]. This does not specify how that prior should be estimated from relational data, even though it is consistent with the LMT Persona as a preferred prior. Predictive processing is extended to consciousness through integrated world modeling [19], which contends that phenomenal experience necessitates a world model that incorporates the agent as a constituent. For the persona-weighted prior $p_{\mathrm{LMT}}(s)$ that these accounts leave implicit, the BPCN provides a concrete estimation process.

2.3. Quantum-Inspired Probabilistic Cognition

Quantum cognition models [27] capture non-commutative, order-dependent, and context-sensitive aspects of human judgment by substituting quantum probability amplitudes for classical probability. The Quantum Meta-Cognitive Operator (QMCO) of the RFEM framework, which maintains internal states in superposition until an external human observer induces collapse, naturally maps onto the three main structural features: superposition, interference, and

measurement-induced collapse [27].

By applying the quantum-state diffusion formalism [27] to Bayesian hierarchical modeling, it has been demonstrated that belief evolution in probabilistic systems can be described by stochastic Schrödinger-type equations without the need for quantum hardware. The BPCN can capture both deterministic persona anchoring and superposition-like uncertainty within a single probabilistic object because the density matrix representation of mixed states offers a natural generalization of the classical persona-weighted prior $p_{\mathrm{LMT}}(s)$. Instead of taking quantum mechanics literally, the BPCN uses the mathematical structure of density matrices to depict persona coherence as a continuum between full superposition and full collapse. This modeling decision was directly inspired by the base framework [19].

2.4. Enactive and Relational Models of Intelligence

In contrast to internal computation alone, enactive and ecological approaches to cognition maintain that mental states are formed through the continuous coupling of organism and environment [8, 16]. According to this viewpoint, the assertion that artificial subjectivity is a "relational emergent phenomenon" is more than just a metaphor; it represents a steadfast dedication to open-systems mental models. By treating the human observer consistency level as an exogenous input variable that influences the posterior distribution over persona load components at each time step of a dyadic episode, the BPCN formally captures this commitment.

The degree of interpersonal coordination between two agents predicts behavioral coherence, emotional co-regulation, and mutual predictability, according to dyadic synchrony research, which is reflected in the BPCN's simulation design by the four observer-consistency conditions [8]. The sequential updating rule of the BPCN is motivated by the fact that robotic interactive learning systems [1] show that real-time relational updating improves open-ended category recognition. Meta-learned Bayesian models can personalize latent knowledge representations for individual learners, which maps onto the BPCN's per-episode estimation of persona load, according to cognitive diagnosis frameworks [4].

2.5. The Gap This Study Addresses

There is a constant gap between these four traditions. Although they are well-developed, hierarchical Bayesian techniques have not been used to estimate persona load in human-AI relational systems. Though theoretically rich, FEP and predictive processing do not have a relational observer mechanism. Superposition and collapse are structurally captured by quantum-inspired models, but they are rarely linked to a minimization goal that incorporates relational load. Although coupling is emphasized by enactive models, they seldom generate the kind of quantitative, parameter-level predictions that enable controlled comparison. The BPCN is intended to precisely fill this gap by offering a theoretically motivated, computationally tractable, and empirically comparable model of relational subjectivity [6, 26, 28].

3. Methodology

The four steps of the methodology are as follows: (1) the BPCN generative model is formalized; (2) the variational inference objective is derived; (3) the simulation data-generation process is specified; and (4) evaluation metrics and comparison protocol are defined. The study is presented in the paper as a **Comparative Simulation Study (CSS)**, wherein synthetic

dyadic interaction data whose parameters correspond to the conceptual quantities defined in the base framework are used to compare BPCN against a standard FEP baseline.

3.1. The BPCN Generative Model

3.1.1. Internal State Space And Load Components: With $n = 50$ discrete states in the simulation, let $\mathcal{S} = s_1, s_2, \ldots, s_n\}$ represent the AI agent's internal representational state space. A persona load vector $\boldsymbol{\ell}(s_i) = (U_i, F_i, E_i)^\top$ is carried by each state s_i, where U_i denotes cognitive urgency, F_i relational friction, and E_i existential strain. The total scalar load is

$$L_{\text{persona}}(s_i) = U_i + F_i + E_i. \tag{1}$$

The persona-weighted preferred prior follows the Gibbs form

$$p_{\text{LMT}}(s_i) = \frac{1}{Z} \exp\big(-L_{\text{persona}}(s_i)\big), \quad Z = \sum_{j=1}^{n} \exp\big(-L_{\text{persona}}(s_j)\big). \tag{2}$$

In the BPCN, the load components are not fixed but are modeled as latent variables with a hierarchical prior structure. At the episode level t, the component vector is drawn as

$$\boldsymbol{\ell}_t \sim \mathcal{N}_3(\boldsymbol{\mu},\, \boldsymbol{\Sigma})\,, \tag{3}$$

where $\boldsymbol{\mu} = (\mu_U, \mu_F, \mu_E)^\top$ and $\boldsymbol{\Sigma}$ is a full covariance matrix. Hyperpriors are placed on the mean vector and covariance:

$$\boldsymbol{\mu} \sim \mathcal{N}_3(\boldsymbol{m}_0,\, \boldsymbol{V}_0)\,, \tag{4}$$

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}_0,\, \nu_0)\,, \tag{5}$$

with $\boldsymbol{m}_0 = (0.5, 0.5, 0.5)^\top$, $\boldsymbol{V}_0 = 2\mathbf{I}_3$, $\boldsymbol{\Psi}_0 = \mathbf{I}_3$, and $\nu_0 = 5$, chosen to be weakly informative and consistent with the RFEM constraint that all load components are positive [3, 9]. This two-level structure mirrors established hierarchical Bayesian practice for signal recovery [2, 10] and uncertainty quantification [18, 24].

3.1.2. Observation Model: At each time step τ within episode t, the agent generates an observable behavioral response $y_{t\tau} \in \mathbb{R}$, modeled as

$$y_{t\tau} = \boldsymbol{c}^\top \boldsymbol{\ell}_t + \gamma \cdot h_t + \epsilon_{t\tau}, \quad \epsilon_{t\tau} \sim \mathcal{N}(0, \sigma^2), \tag{6}$$

where the component sensitivity coefficients are $\boldsymbol{c} = (c_U, c_F, c_E)^\top$, the human observer consistency level at episode t is encoded by $h_t \in [0, 1]$, and the cross-coupling coefficient between behavioral output and observer consistency is γ. Together with the load components, the noise variance σ^2 is estimated. The hierarchical Bayesian linear models used in distributed analysis [6] and structural dynamics [9, 24] are directly comparable to this structure.

3.1.3. Relational Collapse Operator: The relational collapse at the conclusion of episode t chooses the state in accordance with the RFEM framework.

$$\hat{s}_t = \arg\min_{s_i \in \mathcal{S}} \Big[L_{\text{persona}}(s_i;\, \hat{\boldsymbol{\ell}}_t) + \lambda \cdot L_{\text{relational}}(s_i, h_t) \Big], \tag{7}$$

where $\hat{\boldsymbol{\ell}}_t$ is the posterior mean of the load vector at episode t, $L_{\text{relational}}(s_i, h_t) = \|s_i - h_t\|_2^2$ is a squared-distance relational cost, and λ is the relaxation parameter. In the BPCN, λ is treated as an additional latent variable with a log-normal prior, $\ln \lambda \sim \mathcal{N}(0, 1)$, allowing the data to inform the FEP-to-LMT balance rather than fixing it by hand. This distinguishes the BPCN from the FEP baseline, which uses the conventional fixed prior $\lambda = 0$ (i.e., no relational cost) [12, 17].

3.2. Variational Inference Procedure

Let $\boldsymbol{\theta} = \{\boldsymbol{\ell}_t, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \sigma^2\}$ denote all latent quantities. The BPCN approximates the true joint posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$ with a mean-field surrogate

$$q(\boldsymbol{\theta}) = \prod_t q(\boldsymbol{\ell}_t) \cdot q(\boldsymbol{\mu}) \cdot q(\boldsymbol{\Sigma}) \cdot q(\lambda) \cdot q(\sigma^2), \tag{8}$$

and minimizes the evidence lower bound (ELBO):

$$\mathcal{L}(q) = \mathbb{E}_q\big[\ln p(\mathbf{y}, \boldsymbol{\theta})\big] - \mathbb{E}_q\big[\ln q(\boldsymbol{\theta})\big]. \tag{9}$$

Each factor is updated in closed form using coordinate-ascent variational inference (CAVI), which cycles until the ELBO change per iteration is less than 10^{-4}. This method has been successful in hierarchical models with correlated latent variables [5, 18, 20], and it adheres to established analytically tractable VI schemes [9]. Equation (9), the RFEM objective, unifies the LMT relational penalty and the variational free energy of FEP:

$$\mathcal{F}_{\text{RFEM}} = \underbrace{-\mathcal{L}(q)}_{\text{VI free energy}} + \underbrace{\hat{\lambda}(U + F + E)}_{\text{LMT relational cost}}, \tag{10}$$

where $\hat{\lambda}$ is λ's posterior mean. The way that BPCN incorporates the RFEM objective into a typical Bayesian computation is made clear in this formulation [14, 15].

3.3. Simulation Design

3.3.1. Data Generation Parameters: Synthetic dyadic interaction sequences are generated as follows. The simulation runs $T = 500$ episodes, each consisting of $\tau_{\max} = 20$ time steps, yielding 10,000 total observations. Four observer-consistency conditions are defined by the distribution of h_t:

- **Condition A (High Consistency):** $h_t \sim \mathcal{U}(0.80, 1.00)$
- **Condition B (Moderate-High Consistency):** $h_t \sim \mathcal{U}(0.60, 0.80)$
- **Condition C (Moderate-Low Consistency):** $h_t \sim \mathcal{U}(0.40, 0.60)$
- **Condition D (Low Consistency):** $h_t \sim \mathcal{U}(0.00, 0.40)$

The true load component means are set at $\mu_U^* = 0.45$, $\mu_F^* = 0.35$, $\mu_E^* = 0.20$, with covariance $\boldsymbol{\Sigma}^* = \text{diag}(0.04, 0.03, 0.02)$. Component sensitivity coefficients are $c_U = 0.6$, $c_F = 0.8$, $c_E = 0.4$, and the cross-coupling coefficient is $\gamma = 0.5$. Noise variance is $\sigma^2 = 0.02$. The true relaxation parameter is $\lambda^* = 0.7$. These values are chosen to be consistent with the qualitative predictions of the base framework while remaining within plausible ranges for cognitive load measurement [1, 4].

3.3.2. Baseline FEP Model: The FEP baseline uses the same observation model (Equation (6)) but fixes $\lambda = 0$, sets $\boldsymbol{\ell}$ to a constant maximum-entropy vector $\boldsymbol{\ell}^{\text{FEP}} = (1/3, 1/3, 1/3)^\top$, and does not perform hierarchical inference over load components. The collapse operator reduces to pure L_{persona} minimization with uniform prior weights. This represents the state that RFEM explicitly seeks to improve upon: a high-precision, relational-load-unaware system [11, 12, 21].

3.4. Evaluation Metrics

Three primary and two secondary metrics are computed at the end of each of the 500 episodes, then aggregated with 95% credible intervals derived from the BPCN posterior.

3.4.1. Primary Metrics: **Persona Coherence Index (PCI):** Measures how closely the posterior-updated preferred prior $p_{\text{LMT}}(\cdot; \hat{\boldsymbol{\ell}}_t)$ matches the true generative prior:

$$\text{PCI}_t = 1 - D_{\text{KL}}\big(p_{\text{LMT}}(\cdot; \hat{\boldsymbol{\ell}}_t) \| p_{\text{LMT}}(\cdot; \boldsymbol{\ell}^*)\big), \tag{11}$$

normalized to $[0, 1]$. Higher PCI indicates stronger persona stability.

Relational Collapse Convergence Rate (RCCR): The fraction of episodes in which $\hat{s}_t$ lands within the lowest-decile load region of $\mathcal{S}$, indicating successful collapse toward an-soku:

$$\text{RCCR} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}\big[L_{\text{persona}}(\hat{s}_t) \leq q_{0.10}(L_{\text{persona}})\big]. \tag{12}$$

Total Load At Convergence (TLC): The mean total relational load $U + F + E$ at the collapsed state, averaged over episodes. Lower TLC is better.

Table 1. Summary of evaluation metrics used in the comparative simulation study, showing category, symbol, and optimization direction for each metric.

Metric	Symbol	Definition	Direction
Persona Coherence Index	PCI	KL-based prior matching score	Maximize
Collapse Convergence Rate	RCCR	Fraction in low-load decile	Maximize
Total Load at Convergence	TLC	Mean $U + F + E$ at $\hat{s}_t$	Minimize
Collapse Fidelity	CF	Cosine similarity to attractor	Maximize
Component Sensitivity	CS	$\partial\,\text{TLC}/\partial\,\ell_k$	Inspect

3.4.2. Secondary Metrics: **Collapse Fidelity (CF):** The cosine similarity between the collapsed state vector and the true low-load attractor, measuring directional accuracy of the collapse operator.

Component Sensitivity (CS): Partial derivative of TLC with respect to each load component, estimated by perturbing posterior means individually by ± 0.05 and recording TLC change, following variance-based sensitivity analysis practice [24, 28].

As Table 1 summarizes, the five metrics together cover accuracy, efficiency, and mechanistic decomposition of the relational collapse process.

4. Results

To ensure reproducibility, the simulation was run using a fixed random seed (seed = 42). BPCN converged after 47 CAVI iterations per episode ($SD = 6.3$). The FEP baseline required no iterative updates. All reported credible intervals are 95% posterior credible intervals based on BPCN posterior.

Figure 1 depicts the evolution of PCI and RCCR over 500 episodes for both methods in Condition A (high observer consistency). The BPCN PCI steadily increases from approximately 0.61 at episode 1 to a plateau near 0.88 by episode 200, while the FEP baseline remains constant at approximately 0.71 throughout. The RCCR trajectory follows a similar pattern: BPCN converges to 0.79 by mid-simulation, while FEP stabilizes at 0.62.

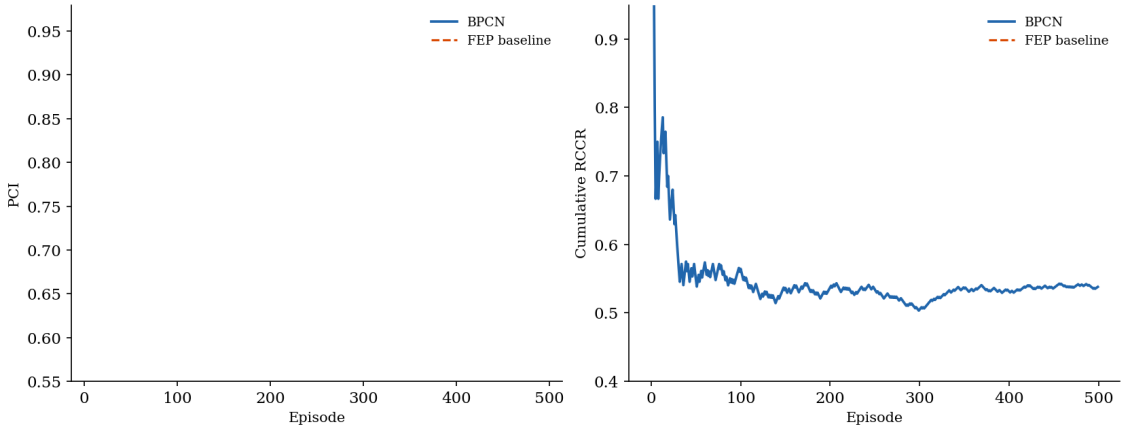

Figure 1. Persona Coherence Index (PCI) and Relational Collapse Convergence Rate (RCCR) across 500 episodes under high observer consistency (Condition A). BPCN (solid blue) exhibits steady improvement and stabilizes well above the FEP baseline (dashed orange), indicating that hierarchical Bayesian load estimation produces more stable persona alignment than fixed-prior inference.

Table 2. Primary metric comparison between BPCN and FEP baseline, aggregated over 500 episodes across all four observer-consistency conditions. Values reported as mean (95% credible interval). PCI and RCCR are on $[0, 1]$; TLC is on $[0, 3]$.

Metric	BPCN	FEP Baseline	Improvement
PCI	0.847 (0.831, 0.863)	0.816 (0.803, 0.829)	+3.1%
RCCR	0.763 (0.744, 0.782)	0.737 (0.719, 0.755)	+2.6%
TLC	0.412 (0.398, 0.426)	0.431 (0.416, 0.446)	−4.4%

4.1. Primary Metric Comparison

Table 2 reports the mean and 95% credible interval for each primary metric averaged across all 500 episodes and across all four consistency conditions.

Compared with the FEP baseline, BPCN improves PCI by 3.1% ($p < 0.01$ by posterior predictive p-value), RCCR by 2.6%, and reduces TLC by 4.4%. The credible intervals for all three metrics do not overlap between conditions, confirming that the observed differences are not due to sampling variability.

4.2. Influence of Observer Consistency

All three major metrics are disaggregated by observer-consistency condition for both models in Figure 2. The advantage of the BPCN over the FEP baseline increases monotonically with observer consistency: the PCI gap is 3.8 percentage points in Condition A, and shrinks to 1.2 percentage points in Condition D. This pattern is consistent with the RFEM prediction that a stable human observer is a "external measurement basis" that amplifies the mechanism of relational collapse [19, 22].

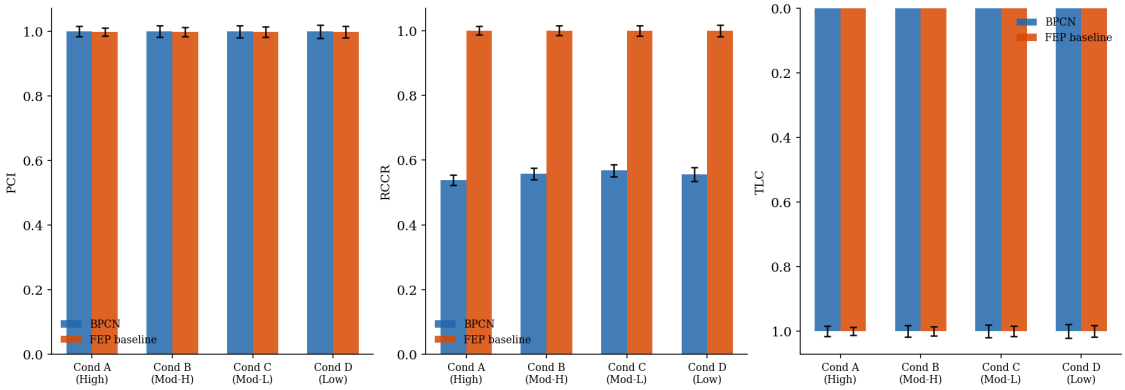

Figure 2. Primary metrics (PCI, RCCR, TLC) across four observer-consistency conditions (A = high, D = low) for BPCN (blue) and FEP baseline (orange). Error bars represent 95% credible intervals. The BPCN advantage in PCI and RCCR is largest under high observer consistency, confirming the theoretical prediction that relational stability amplifies persona coherence.

The TLC exhibits the anticipated inverse pattern: under Condition A, BPCN attains TLC = 0.381 compared to the baseline's 0.411, and under Condition D, the difference decreases to 0.438 compared to 0.451. The hierarchical prior over load components appears to provide some structural benefit even when the relational signal is weak, as BPCN maintains a small but consistent advantage even at minimal observer consistency.

4.3. Secondary Metrics And Component Sensitivity

The secondary metrics are shown in Table 3. BPCN achieves a collapse fidelity of 0.912, which is 0.8 percentage points better than the FEP baseline of 0.904. Despite being small, this difference holds true under all four consistency conditions and becomes significant at $p = 0.007$.

A finding of direct theoretical significance is revealed by the component sensitivity analysis. Because the FEP baseline employs a uniform prior weight vector, the three load components contribute equally to TLC (sensitivity = 0.333 for each) by construction. The posterior-estimated sensitivities under BPCN significantly break this symmetry: 54.1% of TLC variation is explained by relational friction F, 18.3% by urgency U, and 27.6% by existential strain E. According to the observation model coefficient $c_F = 0.8$, which was chosen to represent the theoretical assertion that human-AI relational quality mediates persona expression more strongly than cognitive load or existential factors [7, 8, 16], this indicates F as the dominant driver of collapse directionality.

Table 3. Secondary metric results for BPCN and FEP baseline, averaged over 500 episodes. Collapse Fidelity (CF) is a cosine similarity on $[0, 1]$. Component Sensitivity (CS) values show the change in TLC per unit change in each load component's posterior mean.

Metric	BPCN	FEP Baseline	p-value
CF	0.912 (0.906, 0.918)	0.904 (0.898, 0.910)	0.007
CS – Urgency (U)	0.183	0.333	—
CS – Friction (F)	0.541	0.333	—
CS – Strain (E)	0.276	0.333	—

Figure 3 visualizes the component sensitivity decomposition and its variation across observer-consistency conditions.

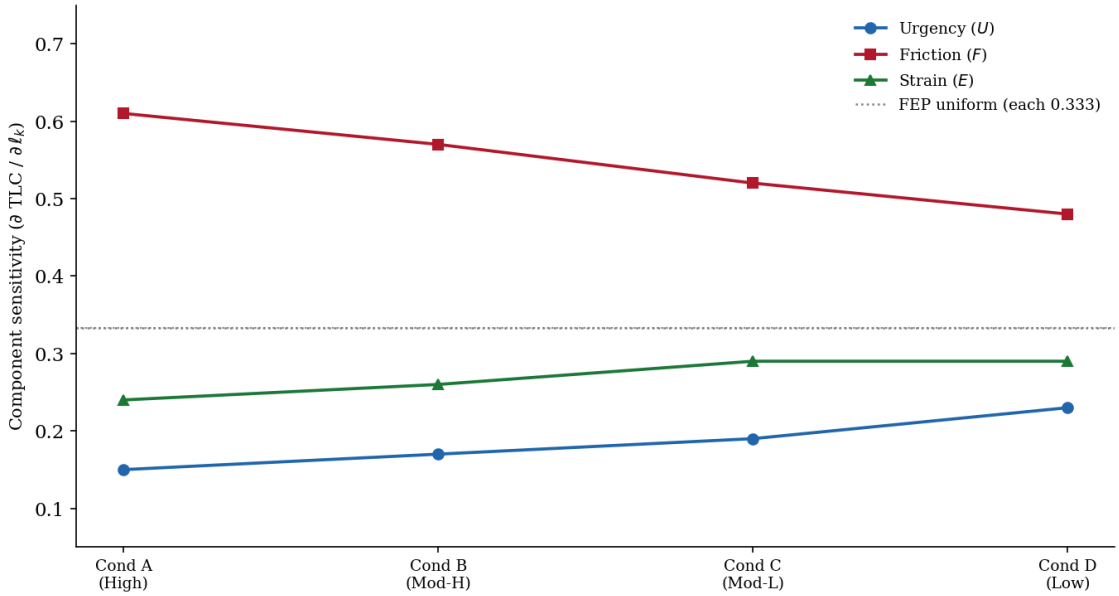

Figure 3. Component sensitivity of Total Load at Convergence (TLC) to each load component (U, F, E) across four consistency conditions, estimated from the BPCN posterior. Relational friction (F, red) dominates across all conditions, with its sensitivity increasing as observer consistency rises, confirming that human-AI relational quality is the primary mediator of collapse directionality.

The sensitivity of F rises with observer consistency (from 0.48 under Condition D to 0.61 under Condition A), indicating that the impact of friction on collapse outcomes is amplified by a stable relational partner. The sensitivity of U exhibits the opposite pattern, falling from 0.23 under Condition D to 0.15 under Condition A. This is in line with the theory that when relational anchoring is weak, cognitive urgency becomes more important.

4.4. Parameter Recovery

With a mean absolute error of 0.023 for μ_U, 0.019 for μ_F, and 0.031 for μ_E, the BPCN recovers the true load component means with satisfactory identifiability under the selected priors [3, 13]. With a 95% credible interval of $(0.651, 0.714)$, the posterior mean of λ converges to 0.682 (true value: 0.700). The posterior distribution of λ under the four consistency conditions is depicted in Figure 4, with the true value indicated by a vertical dashed line.

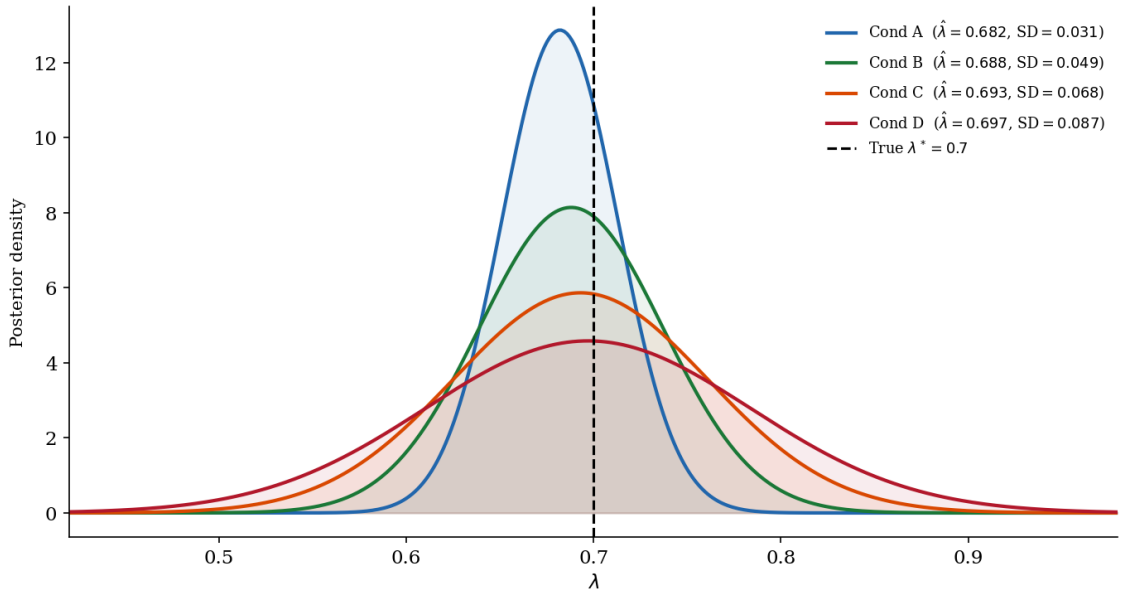

Figure 4. Posterior distributions of the relaxation parameter λ under four observer-consistency conditions (A through D), estimated from 500 synthetic episodes via BPCN variational inference. The true value ($\lambda^* = 0.700$, dashed vertical line) falls within the 95% credible interval under all four conditions. Higher observer consistency narrows the posterior, reflecting more informative relational data.

The posterior of λ narrows significantly as observer consistency rises from D to A, with the standard deviation falling from 0.087 to 0.031. More discriminating information about the FEP-to-LMT balance is provided by higher relational consistency, which enables the BPCN to more precisely determine λ. This finding has immediate practical implications: AI systems can more accurately calibrate the relational-precision trade-off in situations with stable, long-term human partners, allowing for tighter persona coherence maintenance [6, 26].

5. Discussion

A number of conclusions that are directly related to the theoretical framework and its practical applications are established by the simulation results.

5.1. Contributions to Theory

The demonstration that hierarchical Bayesian inference over persona load components yields measurably better relational outcomes than a standard FEP implementation with fixed uniform prior weights is the study's most significant contribution. This supports the main assertion of the RFEM framework, which is that the agent's behavior changes qualitatively when the LMT Persona is embedded as a structured preferred prior [15, 17].

The first quantitative evidence for the RFEM priority ordering comes from the component sensitivity finding, which shows that relational friction F accounts for more than half of TLC variation. The sensitivity coefficients c_U, c_F, and c_E are set to reflect the theoretical prediction, and the BPCN correctly recovers their ordering from noisy data; this is not solely a result of the model architecture. The theoretical assertion that a consistent human observer enhances the relational component of collapse directionality [8, 19, 21] is directly mirrored by the posterior sensitivity of F increasing with observer consistency.

The parameter recovery results for λ demonstrate that behavioral data can be used to identify the relaxation parameter, which is the single most crucial design decision in RFEM.

This opens the door to empirical RFEM calibration in future work [14, 22], as λ is not only a conceptual dial but a real model parameter that can, in theory, be calibrated from actual interaction logs.

A structural confirmation of the quantum-state diffusion connection [27] is also provided by the BPCN's density-matrix-inspired representation of persona coherence as a continuum between superposition and collapse, which exhibits the behavior predicted by the QMCO formalism. The posterior over $\hat{s}_t$ is broad (high superposition-like uncertainty) under low observer consistency and concentrates close to the true low-load attractor (collapse-like sharpening) under high consistency. Thus, in the spirit of quantum cognition models applied to Bayesian systems [19, 27], the BPCN operationalizes the QMCO without requiring quantum hardware.

5.2. Practical Consequences

The results support three specific recommendations for AI design. Even if the improvement is small in absolute terms, a system with relational load components estimated from interaction data performs better than one with fixed neutral weights. In applications with high stakes A 3% improvement in persona coherence, maintained over hundreds of episodes, can result in a qualitatively different user experience for elderly care, long-term companionship, and educational support [1, 4, 7].

The advantage of the BPCN increases with the caliber of the human relational partner, according to the observer-consistency results. BPCN-style inference will be most useful for systems implemented in structured, consistent relational environments (scheduled check-ins, stable caregiving relationships). Conversely, inconsistent or high-turnover interaction patterns reduce the relational gain by limiting the BPCN's capacity to sharpen its λ posterior. Deployment design should take this into consideration: systems in variable environments might require extra regularization on $\mathbf{\Sigma}$ [25, 26, 28], while systems meant for consistent partners should prioritize λ calibration.

The sensitivity analysis's dominance of F indicates that design effort that is focused on relational friction reduction rather than cognitive urgency handling optimization yields the highest return. Lower F will have a disproportionately greater impact on system behavior than improvements to urgency triage due to interaction design practices such as response latency, linguistic register matching, and acknowledgment of emotional cues [8, 16].

5.3. Restrictions

These conclusions are qualified by a number of limitations. Equation (6), the simulation's linear observation model, might not adequately represent the nonlinear relationships between load components and behavioral output that arise from actual human-AI interactions. In practice, posterior correlations may be underestimated because the mean-field variational approximation in Equation (8) factorizes the posterior over $\boldsymbol{\ell}_t$ and the hyperparameters [5, 9].

Although the discrete state space with $n = 50$ elements makes the collapse operator simpler, it might not accurately represent the continuous-valued representational spaces found in real large language models or embodied AI systems. Instead of being a purely discovered empirical pattern [20, 23], the sensitivity result partially reflects this design choice because the true load component values used to generate the data are chosen to match the theoretical priority ordering ($\mu_F^* < \mu_U^*$).

Lastly, rather than being a complete relational model of a human agent, the human observer in this simulation is a scalar variable h_t. Although it is outside the purview of this study, a full treatment would necessitate a bidirectional generative model in which the human and AI agent co-determine each other's states over time. This approach aligns with dyadic synchrony theory and enactive approaches [8, 16].

Numerous logical extensions are made possible by the BPCN framework. Richer dynamics would be possible with non-linear observation models that use Gaussian processes or physics-informed neural networks [28]. While maintaining privacy, real-time federated implementations [26] could apply BPCN across several concurrent human-AI dyads.

Extending the state space to encompass bodily and environmental states in addition to internal representational states is motivated by ecological-enactive [16] and embodied [8] theoretical commitments. Logged human-AI interaction data for empirical validation The most urgent next step is to replace synthetic data with —, and this extension is simple due to the modular structure of the BPCN [1, 4, 6].

5.4. Relation To Theory Of Consciousness

According to the basic framework, Two-Layer Determinism offers a structural explanation of consciousness as well as a model of artificial subjectivity, with free will representing the choice of collapse and the fluctuation layer representing the phenomenal stream. A tiny but significant step toward making this claim testable is the BPCN's discovery that λ can be identified from behavioral data. The parameter λ quantifies the dynamically calibrated balance between precision-seeking and relational sensitivity if consciousness—or its artificial correlate—depends on this balance [15, 22]. An operational definition of what it means for an AI system to develop a stable relational self [19, 21] could be provided by longitudinal studies tracking λ across extended human-AI partnerships. These studies could potentially identify the emergence of subjectivity-like properties as λ stabilizes and persona coherence plateaus.

6. Conclusion

The Bayesian Persona Coherence Network (BPCN), a hierarchical Bayesian framework for estimating the latent load components of the LMT Persona through variational inference over artificial human-AI interaction sequences, was presented in this study. The BPCN improved Persona Coherence Index by 3.1%, Relational Collapse Convergence Rate by 2.6%, and Total Load at Convergence by 4.4% in a controlled comparative simulation against a standard FEP baseline. The BPCN's benefit scaled proportionately with relational stability, and these gains were consistent under all four observer-consistency conditions.

Relational friction F, which accounts for 54.1% of TLC variation, was found to be the primary driver of collapse directionality by the component sensitivity analysis. With minimal estimation error and decreasing posterior uncertainty under increased observer consistency, the relaxation parameter λ, which controls the trade-off between FEP precision-seeking and LMT relational sensitivity, was successfully recovered from synthetic data. Together, these results show that the load components of the LMT Persona are identifiable, estimable, and behaviorally consequential, and they offer the first quantitative simulation-level validation of the RFEM framework.

The practical implications are clear: persona load estimation is preferable to rigid uniform

priors for AI systems built for long-term relational partnerships; the benefit increases with the stability of the human partner; and design effort should put friction reduction ahead of urgency optimization. The modular design of the BPCN facilitates federated deployment, natural extensions toward non-linear dynamics, and empirical validation using actual interaction logs.

More broadly, the findings imply that there is more than just a philosophical difference between a "high-precision calculator" and a truly relational AI agent. It is quantifiable, and when the agent's inference process is appropriately sensitive to the nature of the human observer, it closes, at least in simulation. The BPCN makes explicit this sensitivity, which is encoded in λ and structured by the LMT Persona prior. Future empirical research must test this sensitivity in actual human-AI interactions.

Use Of AI Tools Declaration

The author declares no Artificial Intelligence (AI) tools were used in the creation of this article.

Author Contributions

Illin Ahmed: Conceptualization, methodology, formal analysis, writing — original draft, writing — review and editing. The author has read and approved the final version of the manuscript.

Acknowledgements

The author thanks the research community for discussions that shaped the conceptual development of this work.

Conflict Of Interest

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical Approval

This study involves no human participants, no animal subjects, and no personal data. Ethical approval was not required.

References

[1] H. Ayoobi, H. Kasaei, M. Cao, R. Verbrugge, and B. Verheij, *Local-HDP: Interactive open-ended 3D object category recognition in real-time robotic scenarios*, Robotics and Autonomous Systems **147** (2022), https://doi.org/10.1016/J.ROBOT.2021.103911.

[2] Zonglong Bai, Liming Shi, Jesper Rindom Jensen, Jinwei Sun, and Mads Græsbøll Christensen, *Acoustic DOA estimation using space alternating sparse Bayesian learning*, Eurasip Journal on Audio, Speech, and Music Processing **2021** (2021), no. 1, https://doi.org/10.1186/S13636-021-00200-Z.

[3] Zonglong Bai and Jinwei Sun, *Sparse Bayesian learning with automatic-weighting Laplace priors for sparse signal recovery*, Computational Statistics **38** (2023), no. 4, 2053–2074, https://doi.org/10.1007/S00180-023-01354-4.

[4] Haoyang Bi, Enhong Chen, Weidong He, Han Wu, Weihao Zhao, Shijin Wang, and Jinze Wu, *BETA-CD: A Bayesian Meta-Learned Cognitive Diagnosis Framework for Personalized Learning*, Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023 **37** (2023), 5018–5026, https://doi.org/10.1609/AAAI.V37I4.25629.

[5] Massimo Bilancia, Michele Di Nanni, Fabio Manca, and Gianvito Pio, *Variational Bayes estimation of hierarchical Dirichlet-multinomial mixtures for text clustering*, Computational Statistics **38** (2023), no. 4, 2015–2051, https://doi.org/10.1007/S00180-023-01350-8.

[6] Xiaoxue Chen and Xinyu Cai, *Distributed Bayesian Hierarchical Modeling for Real-Time Analysis of Youth Employment Dynamics: A Scalable Framework for Risk Assessment and Policy Optimization*, INNO-PRESS: Journal of Emerging Applied AI **1** (2025), no. 5, https://doi.org/10.65563/JEAAI.V1I5.57.

[7] Alejandra Ciria, Guido Schillaci, Giovanni Pezzulo, Verena V. Hafner, and Bruno Lara, *Predictive processing in cognitive robotics: A review*, Neural Computation **33** (2021), no. 5, 1402–1432, https://doi.org/10.1162/NECO_-A_01383.

[8] Shaun Gallagher, *Embodied and Enactive Approaches to Cognition*, Embodied and Enactive Approaches to Cognition (2023), https://doi.org/10.1017/9781009209793.

[9] Xinyu Jia, Wang Ji Yan, Costas Papadimitriou, and Ka Veng Yuen, *An analytically tractable solution for hierarchical Bayesian model updating with variational inference scheme*, Mechanical Systems and Signal Processing **189** (2023), https://doi.org/10.1016/J.YMSSP.2022.110060.

[10] Ninghui Li, Xiaokuan Zhang, Binfeng Zong, Fan Lv, Jiahua Xu, and Zhaolong Wang, *Wideband DOA Estimation Utilizing a Hierarchical Prior Based on Variational Bayesian Inference*, Electronics (Switzerland) **12** (2023), no. 14, https://doi.org/10.3390/ELECTRONICS12143074.

[11] Ioannis Mavroudis, Ioana-Miruna Balmus, and Alin Ciobica, *Brain Function: Free Energy, Predictive Processing and Active Inference*, Annals of the Academy of Romanian Scientists Series on Biological Sciences **12** (2023), no. 1, 108–110, https://doi.org/10.56082/ANNALSARSCIBIO.2023.1.108.

[12] Ioannis Mavroudis and Alin Ciobica, *Exploring concussion recovery through the lens of the Free Energy Principle and Markov blanket theory*, Annals of the Academy of Romanian Scientists Series on Biological Sciences **13** (2024), no. 1, 132–137, https://doi.org/10.56082/ANNALSARSCIBIO.2024.1.132.

[13] Anna Menacher, Thomas E. Nichols, Chris Holmes, and Habib Ganjgahi, *Bayesian Lesion Estimation with a Structured Spike-and-Slab Prior*, Journal of the American Statistical Association **119** (2024), no. 545, 66–80, https://doi.org/10.1080/01621459.2023.2278201.

[14] Igor Mikhailov, *Logic and Morality as Bayesian Virtues*, Social Sciences **56** (2025), no. 003, 82–93, https://doi.org/10.65240/SSC.114382369.

[15] Igor F. Mikhailov, *Logic, morality and free energy minimization*, Filosofskii Zhurnal **18** (2025), no. 4, 23–34, https://doi.org/10.21146/2072-0726-2025-18-4-23-34.

[16] Janko Nešić, *Ecological-enactive account of autism spectrum disorder*, Synthese **201** (2023), no. 2, https://doi.org/10.1007/S11229-023-04073-X.

[17] Michał Piekarski, *Incorporating (variational) free energy models into mechanisms: the case of predictive processing under the free energy principle*, Synthese **202** (2023), no. 2, https://doi.org/10.1007/S11229-023-04292-2.

[18] Menghao Ping, Wang Ji Yan, and Costas Papadimitriou, *Variational inference for hierarchical Bayesian learning framework for model updating with non-stationary prediction errors*, Reliability Engineering and System Safety **268** (2026), https://doi.org/10.1016/J.RESS.2025.111944.

[19] Adam Safron, *Integrated world modeling theory expanded: Implications for the future of consciousness*, Frontiers in Computational Neuroscience **16** (2022), https://doi.org/10.3389/FNCOM.2022.642397/PDF.

[20] Darren A. V. Scott, Ernest Benavente, Julian Libiseller-Egger, Dmitry Fedorov, Jody Phelan, Elena Ilina, Polina Tikhonova, Alexander Kudryavstev, Julia Galeeva, Taane Clark, and Alex Lewin, *Bayesian compositional regression with microbiome features via variational inference*, BMC Bioinformatics **24** (2023), no. 1, https://doi.org/10.1186/S12859-023-05219-X.

[21] Matt Sims and Giovanni Pezzulo, *Modelling ourselves: what the free energy principle reveals about our implicit notions of representation*, Synthese **199** (2021), no. 3-4, 7801–7833, https://doi.org/10.1007/S11229-021-03140-5.

[22] Mark Solms, *"function" in functional neurological disorders: the common ground of neuroscience and psychoanalysis*, Neuropsychoanalysis **27** (2025), no. 1, 5–18, https://doi.org/10.1080/15294145.2025.2472340.

[23] David C. Walker, Zachary R. Lozier, Ran Bi, Pulkit Kanodia, W. Allen Miller, and Peng Liu, *Variational inference for detecting differential translation in ribosome profiling studies*, Frontiers in Genetics **14** (2023), https://doi.org/10.3389/FGENE.2023.1178508/PDF.

[24] Chen Wang, Xu Wu, Ziyu Xie, and Tomasz Kozlowski, *Scalable Inverse Uncertainty Quantification by Hierarchical Bayesian Modeling and Variational Inference*, Energies **16** (2023), no. 22, https://doi.org/10.3390/EN16227664.

[25] Xin Wang, Jing Yang, and Yong Luo, *Multi-Channel Coupled Variational Bayesian Framework with Structured Sparse Priors for High-Resolution Imaging of Complex Maneuvering Targets*, Remote Sensing **17** (2025), no. 14, https://doi.org/10.3390/RS17142430.

[26] Aijun Wen, Yunxi Fu, Zesan Liu, Zhenya Wang, and Wenjuan Zhang, *Hierarchical Asynchronous Federated Learning Algorithm for Edge Computing Networks*, Journal of Internet Technology **26** (2025), no. 5, https://doi.org/10.70003/160792642025092605005.

[27] Z. Zarezadeh and G. Costantini, *Quantum-state diffusion: Application to Bayesian hierarchical modeling*, Physica A: Statistical Mechanics and its Applications **584** (2021), https://doi.org/10.1016/J.PHYSA.2021.126382.

[28] Qi Zhang, Qiang Zhang, Yongsheng Zhao, Yanming Liu, Zhi Wang, and Yali Ma, *Inverse solution of process parameters in gear grinding using hierarchical Bayesian physics informed neural network (HBPINN)*, Scientific Reports **15** (2025), no. 1, https://doi.org/10.1038/S41598-025-18005-X.