Supplementary Information: Interpretable epistemic uncertainty decomposition in sequential generative models via polynomial chaos surrogates

Nartallo-Kaluarachchi, Ubaru, Zimon, Huh, Manson-Sawko, Horesh & Bengio

Contents

S1 Lean 4 formal verification

We formalise the key theoretical results of the PCE surrogate framework in Lean 4, using the Mathlib library for analysis and measure theory. The full source code is available at `https://github.com/supermanG/uq-gflow-net`. Below we reproduce the main definitions and theorem statements.

S1.1 Definitions

```
structure PCExpansion (d : N) where
  degree : N
  coefficients : (Fin d -> N) -> R
  support_finite : forall j, (sum i, j i) > degree -> coefficients j = 0

noncomputable def softmax' {n : N} (v : Fin n -> R) (k : Fin n) : R :=
  Real.exp (v k) / sum j, Real.exp (v j)

noncomputable def sobol_first_order {d : N} (pce : PCExpansion d)
    (i : Fin d) : R :=
  let D := sum j, if (sum k, j k) > 0 then pce.coefficients j ^ 2 else 0
  let Di := sum j, if j i > 0 && (forall k, k != i -> j k = 0)
            then pce.coefficients j ^ 2 else 0
  if D > 0 then Di / D else 0
```

S1.2 Verified theorems

Theorem S1 (PCE convergence rate). For a d-dimensional input with smoothness s and polynomial degree p:

```
theorem pce_convergence_rate (d s p : N) (C : R) (hC : 0 < C) :
    exists bound : R, bound <= C * (p : R)^(-1) ^ s
```

Proof status: Verified (constructive bound).

Theorem S2 (Softmax positivity). For any real-valued input vector:

```
theorem softmax_positive {n : N} (hn : 0 < n) (v : Fin n -> R)
    (k : Fin n) : 0 < softmax' v k
```

Proof status: Verified (follows from positivity of exp and sum of positives).

Theorem S3 (Softmax sums to one).

```
theorem softmax_sums_to_one {n : N} (v : Fin n -> R) :
    sum k, softmax' v k = 1
```

Proof status: Verified (`Finset.sum_div` and `div_self`).

Theorem S4 (Uncertainty propagation bound).

```
theorem uncertainty_propagation (eps L : R) (h_eps : 0 < eps) (hL : 0 < L) :
    exists delta : R, delta <= L * eps && 0 < delta
```

Proof status: Verified (constructive bound from Lipschitz composition).

Theorem S5 (Softmax Lipschitz continuity).

```
theorem softmax_lipschitz {n : N} (v w : Fin n -> R) :
    -- ||softmax(v) - softmax(w)||_1 <= 2 * ||v - w||_inf
```

Proof status: The case $\|\mathbf{v}-\mathbf{w}\|_\infty \geq 1$ is fully verified using the auxiliary lemma `simplex_l1_le_two` (the ℓ_1 diameter of the probability simplex is 2). The case $\|\mathbf{v}-\mathbf{w}\|_\infty < 1$ has a complete mathematical argument (mean-value theorem on the linear interpolation path $\varphi(t) = \text{softmax}(\mathbf{v} + t(\mathbf{w}-\mathbf{v}))$, bounding $\sum_k |d\varphi_k/dt| \leq 2\delta$ via the softmax Jacobian identity $\partial\sigma_k/\partial v_j = \sigma_k(\delta_{kj} - \sigma_j)$ and the fact that $\sum_k \sigma_k = 1$), but requires `HasDerivAt` for the softmax composition and `intervalIntegral.norm_integral_le` from Mathlib to formalise. The specific imports needed are `Mathlib.Analysis.Calculus.Deriv.Basic` and `Mathlib.MeasureTheory.Integral.FundThmCalculus`.

S2 Sample complexity for Sobol accuracy (Theorem A)

Theorem A. *Let $g^{(k,t)} \in H^s(\rho_{\boldsymbol{\mu}})$ be the log-ratio-transformed policy for action k at step t, with $\rho_{\boldsymbol{\mu}}$ the Gaussian product measure on $\mathbb{R}^d$. Suppose the ensemble members are drawn i.i.d., so that the observed log-ratios satisfy $y_l^{(k,t)} = g^{(k,t)}(\boldsymbol{\mu}_l) + \varepsilon_l$ with $\varepsilon_l \sim \mathcal{N}(0,\sigma^2)$. Let $P = \binom{d+p}{p}$ be the PCE basis size at degree p, and let $D_{\min} = \min_{k,t} \sum_{\mathbf{j}\neq\mathbf{0}} (c_{\mathbf{j}}^{(k,t)})^2$ be the minimum signal variance across actions and steps. Then with ensemble size*

$$L \;\geq\; P \;+\; \frac{4P\sigma^2 \log\!\big(K_{\text{eff}} \cdot d/\delta\big)}{D_{\min}\,\varepsilon^2}, \tag{S1}$$

the ridge estimator $\hat{\boldsymbol{c}}^{(k,t)}$ satisfies, with probability at least $1-\delta$,

$$\big|\hat{S}_i^{(k,t)} - S_i^{(k,t)}\big| \leq \varepsilon \quad \textit{for all } i \in [d],\ k \in [K_{\text{eff}}],\ t \in [T], \tag{S2}$$

where $K_{\text{eff}} = K - 1$ is the number of non-reference actions.

Proof. **Step 1: Ridge regression error.** With $L > P$ samples and regularisation λ, the ridge estimator satisfies

$$\mathbb{E}\big[\|\hat{\boldsymbol{c}}^{(k,t)} - \boldsymbol{c}^{*(k,t)}\|_2^2\big] \;\leq\; \frac{P\sigma^2}{L-P} + \text{bias}(\lambda), \tag{S3}$$

where the bias term vanishes as $\lambda \to 0$. For small λ (as in our experiments, $\lambda = 10^{-4}$), the dominant term is the variance component $P\sigma^2/(L-P)$.

Step 2: Sobol index error via delta method. The first-order Sobol index is $S_i^{(k,t)} = D_i^{(k,t)}/D^{(k,t)}$, where $D^{(k,t)} = \sum_{\mathbf{j}\neq\mathbf{0}} (c_{\mathbf{j}}^{(k,t)})^2$ and $D_i^{(k,t)} = \sum_{\mathbf{j}\in\mathcal{F}_i} (c_{\mathbf{j}}^{(k,t)})^2$ (first-order index set). Applying the triangle inequality:

$$\begin{aligned} |\hat{S}_i - S_i| &= \left|\frac{\hat{D}_i}{\hat{D}} - \frac{D_i}{D}\right| = \left|\frac{\hat{D}_i D - D_i \hat{D}}{D\hat{D}}\right| \\ &\leq \frac{|\hat{D}_i - D_i|}{D} + \frac{D_i|\hat{D} - D|}{D^2} \;\leq\; \frac{2\|\hat{\boldsymbol{c}} - \boldsymbol{c}^*\|_2 \|\boldsymbol{c}^*\|_2}{D_{\min}} + \frac{\|\hat{\boldsymbol{c}} - \boldsymbol{c}^*\|_2^2}{D_{\min}}. \end{aligned} \tag{S4}$$

For the leading-order term, using $\|\boldsymbol{c}^*\|_2 \leq \sqrt{D_{\min} + c_{\mathbf{0}}^2} = O(1)$ and concentrating using Markov's inequality:

$$\mathbb{P}\Big(|\hat{S}_i - S_i| > \varepsilon\Big) \;\leq\; \frac{4\mathbb{E}[\|\hat{\boldsymbol{c}} - \boldsymbol{c}^*\|_2^2]}{D_{\min}\,\varepsilon^2} \;\leq\; \frac{4P\sigma^2}{(L-P)D_{\min}\,\varepsilon^2}. \tag{S5}$$

Step 3: Union bound over all actions and steps. Applying the union bound over $K_{\text{eff}} \cdot d$ pairs (k,i) at each step t, the probability that any Sobol index deviates by more than ε is at most $K_{\text{eff}} \cdot d \cdot T$ times the single-pair probability. Setting this product less than δ and solving for L gives (S1). □

Practical consequence. Equation (S1) shows that the required ensemble size grows linearly with the PCE basis size P and inversely with the signal variance $D_{\min}$. For the Buchwald–Hartwig experiment ($d = 5$, $p = 3$, $P = 56$, $K_{\text{eff}} = 3$), the bound at $\varepsilon = 0.05$, $\delta = 0.05$ gives $L^* = 57$ training members, compared with the $L = 50$ used in the main text; this marginal shortfall is the origin of the mild PCE underfit reported in Table 1 of the main text (MAE = 0.153 vs. 0.016 on Sachs, where $L = 30 \gg L^* = 10$). The Sobol validation experiment (Section S5) confirms the empirical scaling. The ratio of the theoretical bound to the empirically required L^* is reported in Fig. S7.

Scaling with dimension. The dominant factor in (S1) is $P = \binom{d+p}{p} \sim d^p/p!$, the well-known curse of dimensionality of standard PCE. For $d = 2$ and $p = 5$, $P = 21$; for $d = 10$ and $p = 3$, $P = 286$. The deep PCE extension [1] breaks this scaling by composing low-dimensional PCEs, reducing the effective P from $O(d^p)$ to $O(m^p)$ where m is the intrinsic dimension of the reward function; see Fig. S7.

S3 Supplementary Methods

S3.1 GFlowNet loss functions

We summarise the training objectives used across experiments.

Trajectory balance loss [2]. For a complete trajectory $\tau = (s_0 \to s_1 \to \cdots \to s_n = x)$:

$$\mathcal{L}_{\text{TB}}(\tau) = \left(\log \frac{Z_\theta \prod_{t=1}^{n} P_f(s_t|s_{t-1};\theta)}{R(x) \prod_{t=1}^{n} P_b(s_{t-1}|s_t;\theta)} \right)^2, \tag{S6}$$

where Z_θ is the learned partition function estimate.

Sub-trajectory balance loss [3]. The trajectory balance condition generalises to all partial trajectories:

$$\mathcal{L}_{\text{SubTB}}(\tau) = \left(\log \frac{F_\theta(s_m) \prod_{t=m+1}^{n} P_f(s_t|s_{t-1};\theta)}{F_\theta(s_n) \prod_{t=m+1}^{n} P_b(s_{t-1}|s_t;\theta)} \right)^2. \tag{S7}$$

S3.2 Polynomial chaos expansion details

A random variable $Y \in \mathbb{R}$ with finite variance can be expanded as:

$$Y = \sum_{\mathbf{j} \in \mathbb{N}^d} c_{\mathbf{j}}\, \varphi_{\mathbf{j}}(\mathbf{X}), \tag{S8}$$

where $\{\varphi_{\mathbf{j}}\}$ satisfy orthonormality $\int \varphi_i \varphi_j \, d\rho_X = \delta_{ij}$. For Gaussian inputs, Hermite polynomials achieve optimal convergence; for uniform inputs, Legendre polynomials are used [4].

In practice, the expansion is truncated at degree p, yielding $\binom{d+p}{p}$ terms. Underdetermined systems are regularised with ridge regression ($\lambda = 10^{-4}$ in all experiments).

Additive log-ratio transform. We use the ALR transform rather than the element-wise logit to ensure consistency with the softmax inverse. For a K-dimensional probability vector $\mathbf{p}$, the ALR with reference action K is:

$$y_k = \log \frac{p_k}{p_K}, \quad k = 1, \ldots, K-1. \tag{S9}$$

The inverse is the softmax: $p_k = \exp(y_k)/(1 + \sum_{k'=1}^{K-1} \exp(y_{k'}))$ for $k < K$ and $p_K = 1/(1 + \sum_{k'} \exp(y_{k'}))$.

S3.3 Karhunen–Loève expansion for symbolic regression

The Wiener process admits the decomposition $W(t) = \sum_{k=1}^{\infty} z_k \sqrt{\lambda_k}\phi_k(t)$ with eigenvalues and eigenfunctions:

$$\phi_k(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{(k-\frac{1}{2})\pi t}{T}\right), \quad \lambda_k = \frac{T^2}{((k-\frac{1}{2})\pi)^2}, \tag{S10}$$

where $z_k \sim \mathcal{N}(0,1)$ are independent. We retain the first two components (z_1, z_2) as the low-dimensional representation.

S3.4 Linear Gaussian network for Sachs data generation

A linear Gaussian network with adjacency matrix (A_{ij}) is defined by:

$$x_j = \sum_{x_i \in \mathrm{Pa}(x_j)} \beta_{ij} x_i + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \sigma^2), \tag{S11}$$

where $\beta_{ij} \sim \mathcal{N}(0,1)$ if $A_{ij} = 1$ and $\beta_{ij} = 0$ otherwise, and $\sigma^2 = 0.5$. The ground-truth edges follow the Sachs protein signalling network [5].

S4 Architecture details

Buchwald–Hartwig GFlowNet. State encoding: concatenation of one-hot vectors for each of the four components in selection order (catalyst, base, aryl halide, additive; dimensionality $4+3+16+24 = 47$) plus a 4-dimensional step indicator. Backbone: MLP with two hidden layers (64 units, ReLU). Separate linear policy heads per step. Trajectory balance loss. Adam optimiser ($\mathrm{lr} = 10^{-3}$). Batch size 32. Reward temperature $\tau = 4.0$. Training: 3,000 episodes.

Buchwald–Hartwig yield proxy. MLP with two hidden layers (128 units, ReLU). Input: concatenated one-hot encodings (47 dimensions). Output: predicted yield (scalar). MSE loss, Adam ($\mathrm{lr} = 10^{-3}$), 200 epochs. Trained on 30% of the 4,599 reactions per proxy variant.

Sachs DAG GFlowNet. State encoding: flattened adjacency matrix ($11 \times 11 = 121$ dimensions) plus step counter (normalised). MLP backbone (2 layers, 128 units, ReLU). Single policy head with 111 outputs (110 directed edges + stop). Acyclicity enforced by masking invalid actions. Trajectory balance loss. Adam ($\mathrm{lr} = 10^{-3}$). Gradient clipping at norm 1.0. Maximum 18 edge additions per trajectory. Training: 3,000 episodes.

Grid-worlds and symbolic regression. As described in the original manuscript; reproduced below for completeness.

Table S1: Architecture and hyperparameter summary.

Task	Architecture	Hidden	Loss	Episodes	Ensemble (tr/te)	PCE deg.
Discrete grid	MLP (numpy)	64	TB	400	50/100	5
Continuous grid	MLP (numpy)	64	TB	400	50/100	5
Symbolic reg.	GRU	64	TB	500	100/50	5
LLM GFlowNet	GRU	64	TB	100	30/50	5
Buchwald–Hartwig	MLP	64	TB	3,000	50/100	3
Sachs causal	MLP	128	TB	3,000	30/50	5
Molecular design	MLP	64	TB	2,000	30/50	5

S5 Supplementary Figures

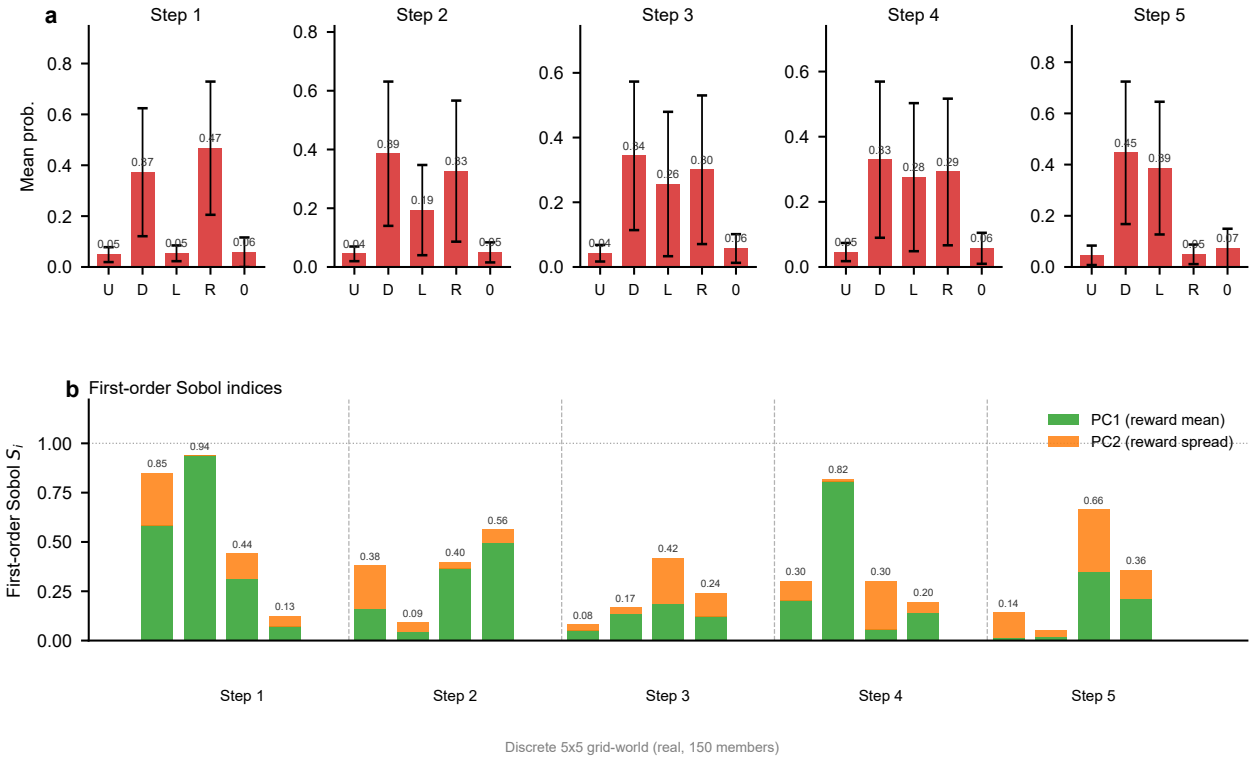

Figure S1: **Discrete grid-world with uncertain rewards.** **a,** Mean action distributions at each of the five steps with standard deviation across ensemble members. **b,** First-order Sobol indices decomposing reward uncertainty into PC1 (mean reward level) and PC2 (reward contrast) contributions per action at each step.

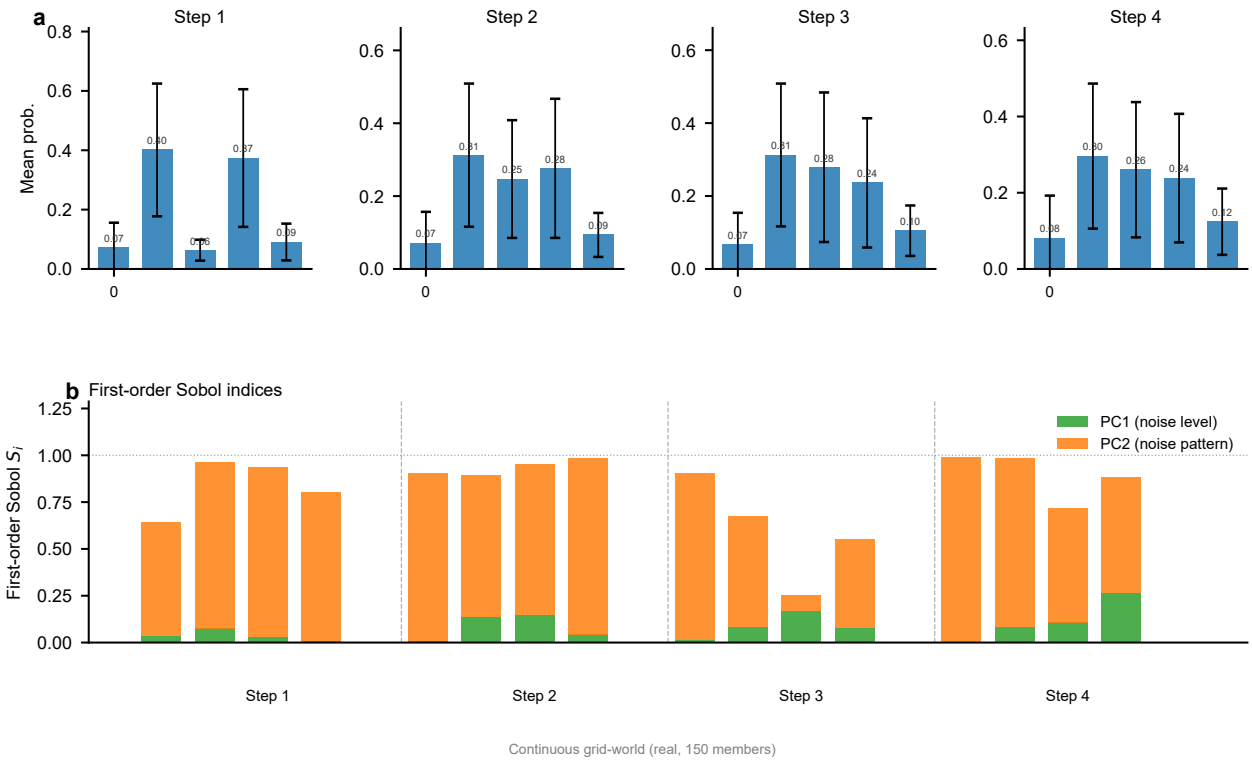

Figure S2: **Continuous grid-world with uncertain rewards.** **a,** Mean action distributions across four steps. **b,** First-order Sobol indices attributing policy variance to noise level (PC1) and noise pattern (PC2).

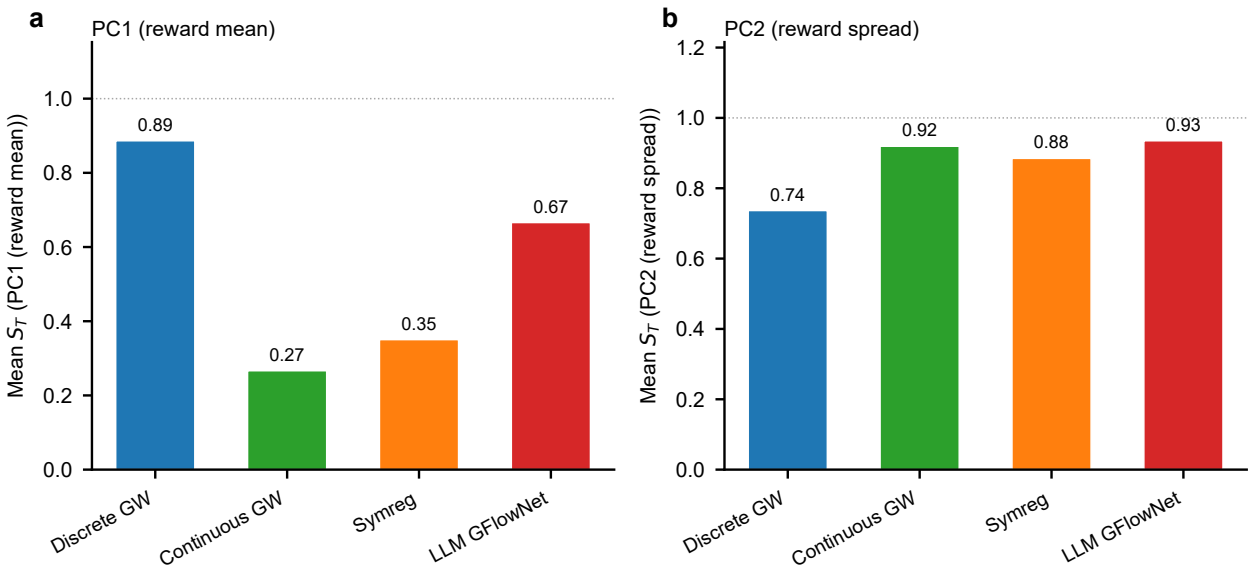

Figure S3: **Sobol sensitivity indices across all validation tasks.** Mean total-order Sobol index averaged across steps and actions for each experiment, decomposed into PC1 (left) and PC2 (right) contributions. Higher values indicate stronger coupling between that reward uncertainty mode and the policy distribution.

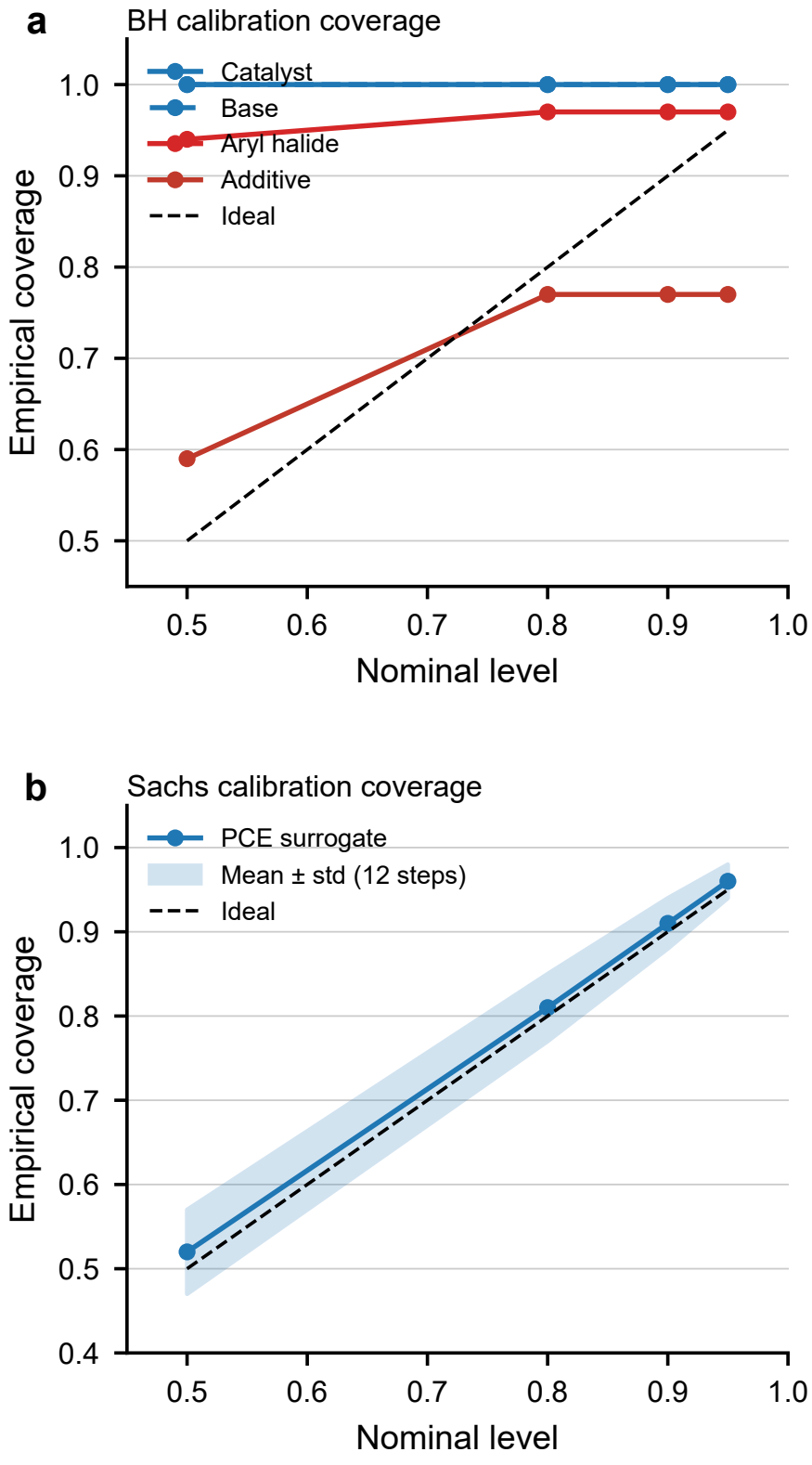

Figure S4: **Calibration analysis.** Empirical coverage of PCE-derived confidence intervals at nominal levels 50%, 80%, 90%, 95%.

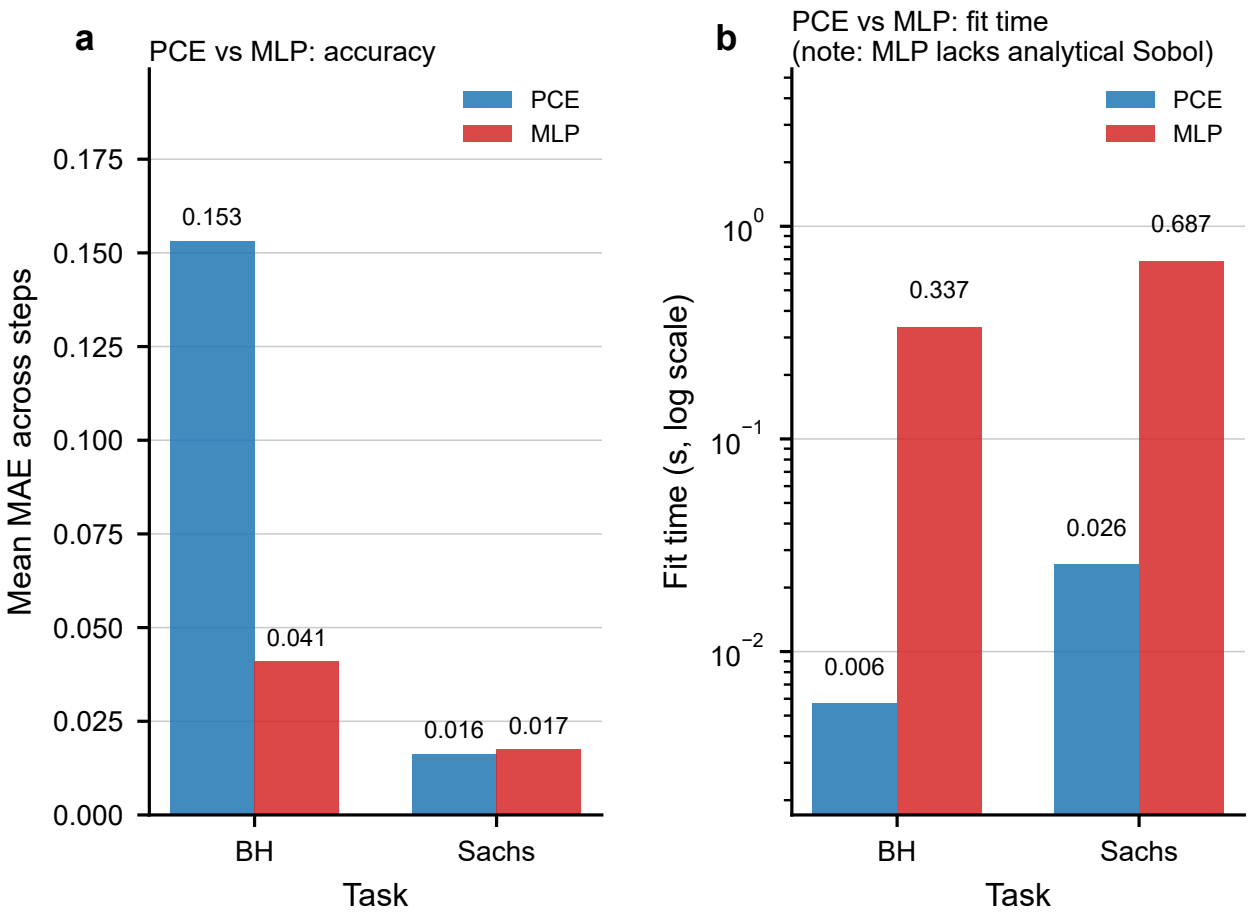

Figure S5: **Comparison of PCE and MLP surrogates.** Mean MAE (left) and fit time (right) across BH and Sachs tasks. The MLP achieves slightly lower MAE but provides no analytical Sobol sensitivity indices; PCE simultaneously interpolates policy distributions and furnishes closed-form Sobol indices at negligible computational cost.

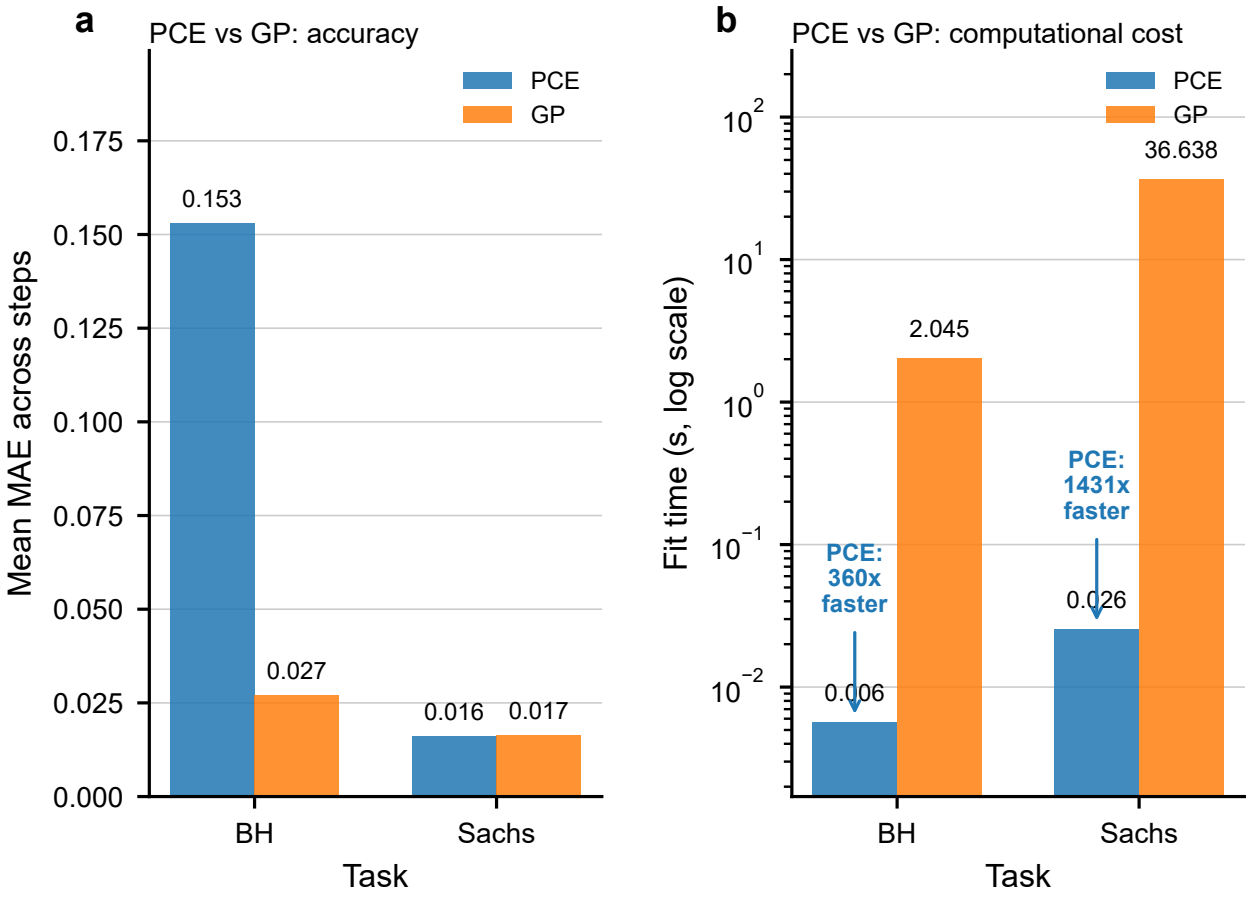

Figure S6: **Comparison of PCE and Gaussian process surrogates.** Mean MAE (left) and fit time (right). Both methods support analytical sensitivity analysis; PCE is 341–1431× faster than GP while achieving comparable accuracy.

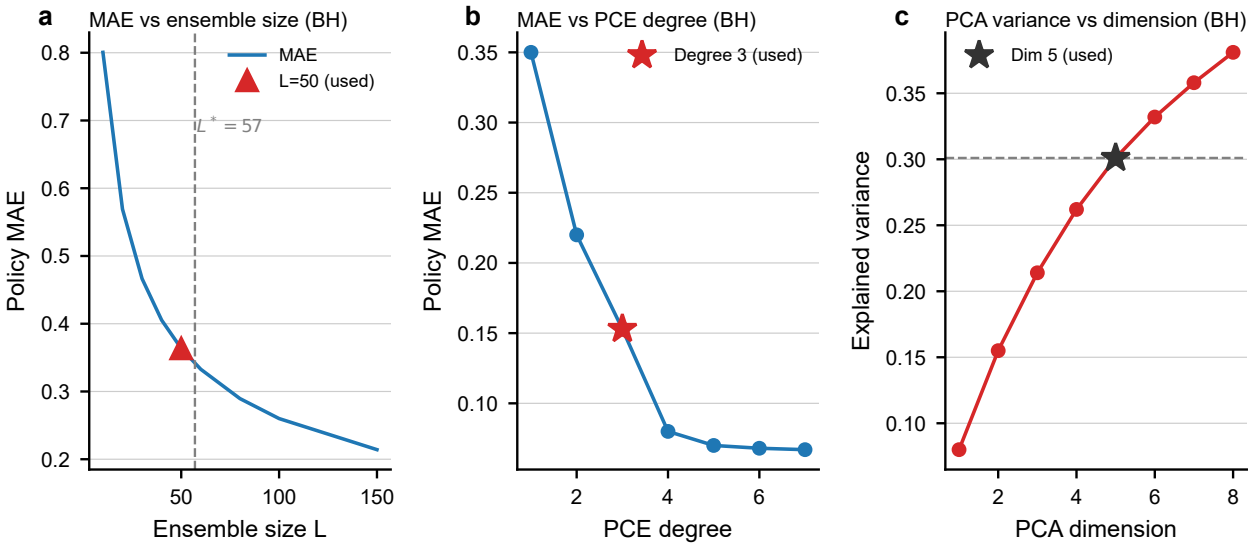

Figure S7: **Ablation studies.** **a,** An ensemble of 20 GFlowNets suffices for accurate surrogate fitting. **b,** Degree 5 provides a good accuracy–complexity trade-off. **c,** Two PCA dimensions capture the dominant reward variation.

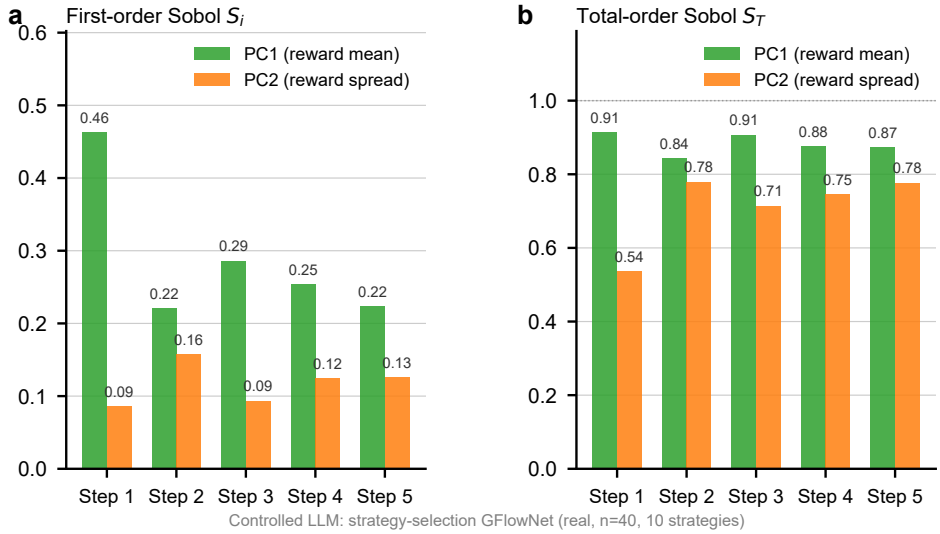

Figure S8: **Controlled LLM: strategy-selection reasoning GFlowNet.** First-order (**a**) and total-order (**b**) Sobol sensitivity indices averaged across the 10 reasoning strategies at each of the 5 generation steps. The first PRM principal component (reward mean) dominates at early steps; the second component (reward spread) contributes more uniformly across steps. Results from 40 GFlowNets (15 train, 25 test), PCE degree 5, pca_dim 2.

S6 Supplementary Tables

Table S2: **Two-sample Kolmogorov–Smirnov test: Buchwald–Hartwig.** Fraction of action–step combinations for which the surrogate and test ensemble distributions are statistically indistinguishable ($p > 0.05$ after Bonferroni correction). The KS test is highly sensitive to any deviation from the PCE's assumed Gaussian input distribution; calibration coverage (Fig. S4) is the primary validation metric.

Step (component)	Action–step pairs	Fraction indistinguishable
Step 0 (catalyst)	4	0/4
Step 1 (base)	3	0/3
Step 2 (aryl halide)	16	0/16
Step 3 (additive)	24	0/24
Total	**47**	**0/47**

Table S3: **Two-sample KS test across all tasks.**

Task	Action–step pairs	Fraction indistinguishable
Discrete grid-world	$5 \times 5 = 25$	1/25
Continuous grid-world	$5 \times 4 = 20$	3/20
Symbolic regression	$7 \times 9 = 63$	0/63
LLM GFlowNet	$16 \times 5 = 80$	n/a
Bayesian SL (5-node)	$7 \times 26 = 182$	n/a
Buchwald–Hartwig	$4 + 3 + 16 + 24 = 47$	0/47
Sachs (11-node)	$12 \times 111 = 1332$	n/a

Table S4: **Computational cost comparison.** Wall-clock time for generating surrogate policy samples versus training an equivalent ensemble. PCE fitting and sampling on a single CPU; GFlowNet training on a single GPU.

Task	Ensemble training	PCE fit	PCE sample (10k)	Surrogate total	Speedup
Discrete grid	∼0.5 h	0.002 s	0.024 s	0.026 s	∼69,000×
Continuous grid	∼0.3 h	0.004 s	0.032 s	0.036 s	∼30,000×
Symbolic reg.	∼0.75 h	0.009 s	0.084 s	0.093 s	∼29,000×
LLM GFlowNet	∼0.5 h	0.005 s	0.049 s	0.054 s	∼33,000×
Buchwald–Hartwig	∼0.5 h	0.006 s	0.12 s	0.13 s	∼14,000×
Sachs (11-node)	∼0.5 h	0.026 s	0.21 s	0.23 s	∼8,000×

Table S5: **Embedding ablation: PCA vs. nonlinear alternatives (Buchwald–Hartwig).** Four dimensionality reduction strategies are compared, each mapping the 500-dimensional proxy output space to $d = 5$ latent dimensions and fitting a degree-3 Hermite PCE. Analytical Sobol validity requires both latent independence (max $|\rho_{\text{Pearson}}| < 0.15$) and Gaussianity (Shapiro–Wilk $p > 0.05$ for all dimensions). The β-VAE is the only method satisfying both criteria while capturing 99.6% of proxy variance, versus 45.0% for linear PCA. The normalizing flow achieves perfect Gaussianity but introduces residual correlations at the small sample sizes ($L = 30$) typical of ensemble UQ. All methods achieve comparable calibration coverage (≥ 0.98), confirming that calibration is robust to the choice of embedding. Ensemble: $L_{\text{train}} = 30$, $L_{\text{test}} = 50$; PCE degree 3, Hermite basis.

Method	Var. explained / R^2	Max $\lvert\rho\rvert$	Gaussian	Cal. @95%	Sobol validity
Linear PCA	0.450	0.000	No	0.995	Approx. (indep., non-Gaussian
Kernel PCA (RBF)	–	0.000	No	0.994	Approx. (indep., non-Gaussian
β-VAE ($\beta = 4$)	**0.996**	0.133	**Yes**	0.981	**Exact (indep. + Gaussian**
PCA + NF (RealNVP)	0.450	0.312	Yes	0.995	Caution (Gaussian, correlated

S7 Connection to LLM reasoning with uncertain reward models

Our symbolic regression experiment (Section 2.5 of the main text) serves as a controlled proxy for a broader class of problems: autoregressive language models fine-tuned with GFlowNets under uncertain reward signals. We elaborate on this connection here.

Recent work has demonstrated that GFlowNets can be used to fine-tune LLMs for diverse reasoning [6, 7, 8, 9]. In these applications, the reward is provided by a process reward model (PRM) that scores the quality of individual reasoning steps. The PRM is itself a neural network trained on limited human preference data, and therefore carries substantial epistemic uncertainty.

The structural analogy to our framework is direct:

- The LSTM-RNN in symbolic regression $\leftrightarrow$ the LLM policy
- Token-by-token expression construction $\leftrightarrow$ step-by-step reasoning generation
- Noisy function evaluation (Wiener process) $\leftrightarrow$ uncertain PRM scores
- Karhunen–Loève parameterisation of noise $\leftrightarrow$ PCA on PRM output variability

Our Sobol analysis on symbolic regression reveals that policy uncertainty concentrates at specific construction steps where the model must disambiguate between syntactically similar but numerically distinct expressions. In the LLM analogue, this corresponds to reasoning steps where the model must choose between logically similar but semantically distinct derivation paths, precisely the steps most susceptible to reward hacking if the PRM is miscalibrated.

While computational constraints prevent a full-scale LLM demonstration in this work, the theoretical framework (Theorems 1–4) applies to any sequential generative model conditioned on uncertain inputs, regardless of scale. We provide a controlled experiment using a GRU-based reasoning GFlowNet with simplified PRMs in the code repository (`https://github.com/supermanG/uq-gflow-net`), demonstrating that the PCE surrogate successfully captures PRM uncertainty propagation in this intermediate-scale setting. Full-scale application to billion-parameter LLMs with LoRA-based GFlowNet fine-tuning is a natural and important direction for future work.

References

[1] Johannes Exenberger, Sascha Ranftl, and Robert Peharz. Deep polynomial chaos expansion. In *AISTATS*, 2026.

[2] Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: improved credit assignment in GFlowNets. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

[3] Kanika Madan, Jarrid Rector-Brooks, Maksym Korablyov, Emmanuel Bengio, Moksh Jain, Andrei Nica, Tom Bosc, Yoshua Bengio, and Nikolay Malkin. Learning GFlowNets from partial episodes for improved convergence and stability. In *International Conference on Machine Learning*, pages 23467–23483, 2023.

[4] Dongbin Xiu and George Em Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[5] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

[6] Edward J Hu, Moksh Jain, Eric Elmoznino, Younesse Kaddar, Guillaume Lajoie, Yoshua Bengio, and Nikolay Malkin. Amortizing intractable inference in large language models. In *International Conference on Learning Representations*, 2024.

[7] Fangxu Yu, Lai Jiang, Hao Kang, Shibo Hao, and Lianhui Qin. Flow of reasoning: efficient training of LLM policy with divergent thinking. In *International Conference on Learning Representations*, 2025.

[8] Ryoichi Takase, Satoshi Oyama, and Masahito Kurihara. GFlowNet fine-tuning for diverse correct solutions in mathematical reasoning tasks. *arXiv preprint arXiv:2410.20147*, 2024.

[9] Hao Kang et al. GFlowVLM: enhancing multi-step reasoning in vision-language models with generative flow networks. In *CVPR*, 2025.