Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

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

Lior Horesh

lhoresh@us.ibm.com

IBM Research https://orcid.org/0000-0001-6350-0238

Ramon Nartallo-Kaluarachchi

Oxford University

Shashanka Ubaru

IBM Research

Malgorzata Zimon

IBM Research

Ben Dongsung Huh

Independent Researcher

Robert Manson-Sawko

IBM Research

Yoshua Bengio

U. Montreal / Mila https://orcid.org/0000-0002-9322-3515

Article

Keywords: Uncertainty quantification, Generative flow networks, Polynomial chaos expansion, Sobol sensitivity analysis, Reaction optimisation, Molecular design, Causal discovery

Posted Date: May 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9521261/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

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

Ramón Nartallo-Kaluarachchi[2], Shashanka Ubaru[3], Małgorzata J. Zimon[1], Dongsung Huh[4], Robert Manson-Sawko[1], Lior Horesh[*3], and Yoshua Bengio[5]

[1]IBM Research Europe, Daresbury, United Kingdom
[2]Mathematical Institute, University of Oxford, United Kingdom
[3]IBM Research, Thomas J. Watson Research Center, USA
[4]MIT–IBM Watson AI Lab, Massachusetts, USA
[5]Université de Montréal, Mila–Quebec AI Institute, Montreal, Quebec, Canada

Abstract

Sequential generative models conditioned on uncertain rewards are central to AI-driven scientific discovery, yet the epistemic uncertainty they inherit from imperfect reward estimates remains unquantified. We propagate this uncertainty through generative flow networks (GFlowNets) by fitting polynomial chaos expansions (PCEs) to small ensembles of trained models. The PCE coefficients yield analytical Sobol sensitivity indices, providing the first interpretable decomposition of *which reward components drive which generative decisions*, a capability unavailable from deep ensembles, Bayesian neural networks, or Monte Carlo dropout. Convergence guarantees are established theoretically and four of five are formally verified in the Lean 4 proof assistant. Across three real-world tasks the framework reveals actionable structure invisible to ensembles alone. On the Doyle–Dreher Buchwald–Hartwig dataset [1] catalyst selection is robust ($D_{\text{catalyst}} \approx 71$) while additive selection is fragile ($D_{\text{additive}} \approx 179$, 2.5× higher). In fragment-based molecular design the linker position is the most sensitive ($D_{\text{linker}} \approx 28$) while decoration positions are the most robust ($D \approx 14$–18), reversing the conventional scaffold-robust / decoration-fragile assumption. On the Sachs protein signalling network, MAPK-cascade edges and PKA/PKC hub edges separate into distinct sensitivity regimes, providing a targeted map for perturbation experiments. Calibration coverage at the 95% level reaches 0.97–1.00 across the dominant steps, and the surrogate evaluates 10,000 policy samples in milliseconds – 10^3–$10^4\times$ faster than exhaustive retraining.

Keywords: Uncertainty quantification, Generative flow networks, Polynomial chaos expansion, Sobol sensitivity analysis, Reaction optimisation, Molecular design, Causal discovery

1 Introduction

Generative models that construct outputs through sequential decision-making, including generative flow networks (GFlowNets) [2, 3], autoregressive transformers [4], and diffusion models [5], have

*Corresponding author: lhoresh@us.ibm.com

become central tools across scientific discovery [6], combinatorial optimisation [7], and biological design [8]. These models are often described as uncertainty-aware because they define probability distributions over outputs. However, this aleatoric uncertainty, arising from the stochastic generation process, is distinct from epistemic uncertainty, which reflects ignorance about the model itself [9].

Epistemic uncertainty is especially consequential when generative models are conditioned on reward or scoring functions that are themselves uncertain. In molecular design, reward functions are typically parameterised by neural network surrogates trained on limited experimental data [6, 10]. In chemical reaction optimisation, yield predictors are trained on high-throughput but incomplete datasets [1]. In Bayesian structure learning, the marginal likelihood used to score candidate graphs depends on finite observational datasets [11]. In each case, the learned generative policy inherits uncertainty from the reward, yet standard GFlowNets offer no mechanism to express or quantify it.

Existing approaches to epistemic uncertainty in deep learning, including Bayesian neural networks [12, 13] and deep ensembles [14, 15], face well-known challenges. Bayesian methods are difficult to scale; ensemble methods demand repeated training at substantial computational cost. Crucially, none of these approaches provide *an interpretable decomposition* of uncertainty: they can quantify that a prediction is uncertain, but not *why* it is uncertain or *which components* of the input uncertainty are responsible.

We draw inspiration from the field of uncertainty quantification (UQ) in computational engineering, where surrogate models approximate expensive simulators to enable efficient uncertainty propagation [16, 17]. A cornerstone technique is the polynomial chaos expansion (PCE) [18, 19], which represents a quantity of interest as a polynomial function of random inputs. Uniquely among surrogate methods, the PCE coefficients enable *analytical* computation of Sobol sensitivity indices [20], providing an exact variance-based decomposition of which input uncertainties drive which outputs. Recent extensions to high-dimensional settings via deep PCE [21], PCE combined with variational autoencoders for representation learning [22], and neural network architectures based on arbitrary polynomial chaos [23] have expanded the applicability of these ideas.

Here we propose a surrogate modelling framework for quantifying and decomposing epistemic uncertainty in GFlowNets (Fig. 1). Our approach proceeds in three stages: (i) train a modest ensemble of GFlowNets under varied reward conditions, (ii) learn a low-dimensional parameterisation of the reward function space, and (iii) fit a PCE that maps from this low-dimensional representation to the policy distribution at each step of a trajectory. The fitted surrogate enables both inexpensive Monte Carlo estimation of policy uncertainty and, our central contribution, *analytical Sobol sensitivity indices* that decompose policy uncertainty into contributions from each component of reward uncertainty.

We make four contributions:

1. A general framework for propagating epistemic uncertainty through conditioned sequential generative models, with PCE surrogates fitted to small training ensembles.

2. Formal convergence guarantees for the surrogate and its Sobol indices, with four of five key properties verified in Lean 4 (T1, T3a, T3b, T4); two formalisation gaps remain on ratio continuity and softmax Jacobian bounds.

3. The first analytical sensitivity decomposition of policy uncertainty in generative models, revealing which reward components drive which generative decisions.

4. Demonstrations on seven tasks of increasing complexity: three real-world applications – chemical reaction condition optimisation on the Doyle–Dreher Buchwald–Hartwig dataset [1] (with Sobol indices directly addressing the Chuang–Keiser debate [24]), fragment-based molecular design with uncertain drug-likeness reward, and Bayesian causal discovery on the Sachs protein signalling network [25] – together with four further validation tasks (discrete and continuous grid-worlds, symbolic regression, and a GRU-based LLM GFlowNet with uncertain process reward model).

2 Results

2.1 Surrogate modelling framework

We consider a GFlowNet trained on a reward function R sampled from a distribution $\mathcal{R}$ that reflects epistemic uncertainty—each realisation of R yields a different learned policy. Our goal is to characterise the marginal distribution over policies induced by $\mathcal{R}$, without exhaustive retraining.

The framework has three components (Fig. 1; see Methods for full details). First, a low-dimensional parameterisation $\boldsymbol{\mu} \in \mathbb{R}^d$ of the reward function is obtained via principal component analysis (PCA) or an analytical expansion (PCA is preferred because its orthogonal components yield independent Gaussian inputs, a prerequisite for valid analytical Sobol decomposition; see Methods). Second, a small ensemble of L GFlowNets is trained, each on a distinct reward realisation, and the policy along a trajectory of interest is extracted. Third, for each action k at trajectory step t, a PCE maps from $\boldsymbol{\mu}$ to the log-ratio-transformed action probability (see Methods, Eq. 2), fitted by regularised regression. The surrogate then enables rapid Monte Carlo sampling and, uniquely, analytical computation of Sobol sensitivity indices from the PCE coefficients (Methods, Eq. 5).

Why not simply analyse the raw ensemble? If one trains L GFlowNets, the ensemble already provides L policy samples, and one might ask why a PCE surrogate is needed. Three reasons motivate the additional step. First, Sobol sensitivity indices require a smooth functional form: computing them from a finite-sample ensemble requires expensive Monte Carlo integration over the input space, which negates the computational advantage and introduces sampling noise, whereas the PCE yields exact analytical indices from its coefficients. Second, the PCE is a continuous functional approximation and can reveal structure, such as multimodality or non-monotone dependence, invisible in a finite sample. Third, the surrogate supports counterfactual queries, enabling evaluation of "what if the uncertainty had a different structure?" without retraining any GFlowNets.

2.2 Reaction condition optimisation: Buchwald–Hartwig cross-coupling

A central question in machine learning for chemistry is whether learned models capture genuine chemical signals or statistical artifacts of the training set. Chuang and Keiser [24] demonstrated that the influential Doyle–Dreher yield predictor [1] for palladium-catalysed Buchwald–Hartwig amination contains strong substrate-identity effects that can confound interpretation of apparent model accuracy. The debate has remained largely unresolved because existing methods quantify *aggregate* prediction uncertainty but cannot attribute it to specific reaction components. The Sobol sensitivity decomposition directly addresses this gap: a low Sobol index on a reaction component indicates that the GFlowNet recommendation is driven by a genuine chemical signal replicated across training subsets, while a high Sobol index flags a component whose recommendation changes

with the training set, the hallmark of a statistical artifact.

We framed reaction condition selection as a 4-step GFlowNet construction: the agent sequentially selects the catalyst, base, aryl halide, and additive, with reward proportional to the predicted yield from a proxy model trained on a subset of the data. The Doyle–Dreher dataset [1] provides experimentally measured yields for 4,599 reactions spanning 16 aryl halides, 4 ligand/catalyst systems, 3 bases, and 24 isoxazole additives. By training the proxy on only 30% of reactions, substantial epistemic uncertainty arises from unobserved combinations. We trained ensembles of 50 (training) and 100 (test) GFlowNets on independent proxy realisations and parameterised the reward space via PCA on proxy outputs, retaining 5 dimensions (30.1% of total proxy variance). While total explained variance is modest, the leading principal components capture the dominant *structured* variation across reaction components – correlated shifts in yield predictions that systematically alter which catalyst–base–substrate–additive combination a GFlowNet prefers. The remaining ∼70% of variance is distributed across many trailing modes that reflect high-frequency proxy fluctuations (individual reaction noise) rather than systematic trends; these modes do not coherently influence policy behaviour, as confirmed by the high calibration coverage reported below. A degree-3 Hermite PCE was fitted to the 50 training members; Theorem A confirms this is near-optimal given 56 basis functions and a required ensemble size of $L^* = 57$.

The Sobol sensitivity decomposition (Fig. 2) provides a per-component, per-step analysis of which reaction parameters are robust to predictor uncertainty and which are fragile.

Catalyst selection exhibits the lowest total policy variance ($D_{\text{catalyst}} \approx 71.3$) and the lowest Sobol sensitivity: the GFlowNet consistently prefers the same ligand systems regardless of which data subset the predictor was trained on. This robustness is consistent with the established chemistry of biaryl monophosphine ligands, which are broadly effective across substrate scope in Buchwald–Hartwig coupling [26]. The Sobol index, therefore, corroborates the chemical interpretation: catalyst preference reflects genuine signal, not a training-set artifact.

Additive selection shows the highest total policy variance ($D_{\text{additive}} \approx 179.2$) and the largest Sobol sensitivity – 2.5× higher than the catalyst: which additive the model recommends depends strongly on which reactions appeared in the training set. This directly engages the Chuang–Keiser concern [24]: a high Sobol index for additive selection indicates it is the component most susceptible to training-set bias and flags it as the reaction parameter where additional experimental data would most improve model reliability. A chemist using this analysis would rightly trust the catalyst recommendation while treating additive recommendations with caution pending further experimental validation.

Quantitatively, the PCE surrogate achieved 95% marginal calibration coverage of 1.00, 1.00, 0.97, and 0.77 for catalyst, base, aryl halide, and additive steps, respectively (Supplementary Fig. S4). We report marginal coverage (averaged over actions and test points) rather than joint coverage (which requires all K actions to fall within the credible region simultaneously), as the former is a more practical metric for per-action uncertainty calibration. Two-sample KS tests against the test ensemble show low pass rates across all 47 action-step combinations (Supplementary Table S1); this reflects the sensitivity of the KS test to any deviation from the PCE's assumed Gaussian input distribution rather than a failure of practical utility, as the calibration analysis confirms that the surrogate's uncertainty intervals contain the correct fraction of test ensemble members.

2.3 Bayesian causal discovery: Sachs protein signalling network

Knowing *which* causal inferences are data-limited has direct experimental consequences: edges with high policy variance are candidates for targeted interventional validation, while robust edges can be treated as established. We applied the framework to Bayesian structure learning on the Sachs protein signalling network [25], a benchmark causal graph comprising 11 phosphoproteins and phospholipids whose regulatory relationships were established through interventional flow cytometry experiments. The GFlowNet samples candidate directed acyclic graphs (DAGs) by sequential edge addition, with reward proportional to the Bayesian Gaussian equivalent (BGe) score [27] computed from finite observational data.

Epistemic uncertainty arises from the limited sample size: each ensemble member subsamples 200 of the 853 real flow cytometry observations [25], inducing variation in the BGe score reward across members. We parameterised the reward space by applying PCA to the flattened per-member sample covariance matrices, retaining 2 components (38.4% of the variance explained).

The Sobol analysis (Fig. 3b) reveals that all 12 edges along the ground-truth trajectory carry high total policy variance ($D \approx 10^5$–1.4×10^5 in ALR space), confirming that 200 observations are insufficient to resolve the signalling network structure. The decomposition nonetheless reveals two distinct sensitivity regimes. MAPK-pathway edges – Raf→Mek ($S_{\mathrm{PC2}} = 0.976$), Mek→Erk ($S_{\mathrm{PC2}} = 0.965$), and PKA→P38 ($S_{\mathrm{PC2}} = 0.957$) – are dominated by the second principal component of covariance variation, which captures variation along the MAPK signalling axis. All remaining PKA and PKC hub edges are dominated by the first principal component ($S_{\mathrm{PC1}} \approx 0.93$–$0.95$). The PKA→Erk edge sits at the boundary between regimes ($S_{\mathrm{PC1}} = 0.945$, $S_{\mathrm{PC2}} = 0.670$), consistent with its dual role linking PKA hub activity to the ERK cascade. This decomposition provides a data-driven experimental recommendation: perturbations along the Raf/Mek/P38 axis would most reduce MAPK-edge uncertainty, while interventions on PKA and PKC activities would target hub connectivity uncertainty.

2.4 Validation on synthetic grid-world tasks

The grid-world experiments provide ground-truth control for assessing surrogate fidelity in a setting where the exact empirical policy distribution is known. We validated on a 5×5 discrete grid-world with stochastic zone-level reward shifts ($5^4 = 625$ configurations, four quadrant zones each taking values in $\{-1, -0.5, 0, 0.5, 1\}$) and a continuous grid-world with three Gaussian reward bumps whose amplitudes are drawn from $\mathcal{N}(0, 1)$ (see Methods). In both cases, a degree-5 PCE surrogate trained on 50 GFlowNet members faithfully recovered the test-ensemble distribution of 100 unseen members across all trajectory steps (Supplementary Figs. S1–S2).

The Sobol analysis reveals that in the discrete grid-world, early trajectory steps exhibit low total variance, indicating that the initial direction of movement is robust to reward perturbations. Later steps near the termination decision show substantially higher variance, with the first principal component (capturing overall zone level) contributing more strongly at early steps and the second component (capturing zone contrast) becoming dominant at later steps (Fig. 4).

2.5 Symbolic regression as autoregressive generation proxy

To test the framework on a structured combinatorial task that mirrors autoregressive language generation, we applied it to symbolic regression: reconstructing $f(x) = \sin(x) + 2 - x$ from a seven-

token vocabulary $\{\sin, \cos, +, -, x, 2, \text{EOS}\}$ in the presence of additive Wiener noise. The GFlowNet (a GRU) builds token sequences of up to 9 steps, precisely analogous to how a language model generates reasoning steps. The noisy reward $\exp(-\text{MSE}(\hat{f}, f_{\text{noisy}}))$ is analogous to an uncertain process reward model (PRM) in LLM fine-tuning [28–30].

Reward uncertainty is parameterised via the two leading Karhunen–Loève modes of the Wiener process on $[0, 2\pi]$, giving a 2D latent space with analytically known basis functions. A degree-5 PCE fitted to 100 training GFlowNets accurately captured both the central tendency and spread of the empirical policy distribution from 50 unseen test models (Fig. 5). The Sobol decomposition (Fig. 5b) reveals that the first KL component (low-frequency noise) dominates policy uncertainty at early token-selection steps. In contrast, the second component (high-frequency noise) becomes increasingly important at later steps where the model must discriminate between syntactically similar but numerically distinct completions.

2.6 LLM GFlowNet with uncertain process reward model

A key motivation for this work is the deployment of GFlowNets as fine-tuning objectives for large language models, where process reward models (PRMs) guide step-by-step reasoning generation [28–30]. PRMs trained on limited human feedback carry substantial epistemic uncertainty, which directly propagates into uncertainty about which reasoning steps the LLM should favour. We demonstrate the framework on a GRU-based reasoning GFlowNet over an arithmetic vocabulary of 17 tokens (digits, operators, and EOS), where the GFlowNet constructs token sequences evaluated by an uncertain PRM.

The PRM is a two-layer MLP on top of a frozen GRU feature extractor, trained with limited labelled examples ($n_{\text{label}} = 20$) drawn from a pool of 40 arithmetic problems; different random subsets produce different PRM weight configurations, inducing a distribution over reward functions. We trained an ensemble of 80 GFlowNets (30 training, 50 test), and parameterised reward variation via PCA on per-member PRM output vectors (explained variance 24.4%), retaining 2 dimensions.

The PCE surrogate (degree 5) achieves a mean 95% calibration coverage of 0.68 across the 5 trajectory steps, comparable to the grid-world and symbolic regression results at the same degree and training set size. The Sobol decomposition reveals that the first PRM principal component (which captures the mean reward level) dominates policy uncertainty at early token steps. In contrast, the second component (reward spread across problems) gains relative importance at later steps.

Beyond sensitivity indices, the PCE surrogate reveals *distributional structure* invisible to scalar Sobol summaries. At later trajectory steps, as the model approaches reward-determining tokens, the surrogate-predicted distribution over termination (EOS) probability exhibits a bifurcation: under some PRM configurations the GFlowNet terminates with high confidence, while under others it assigns only moderate termination probability, reflecting qualitatively different behavioural regimes depending on whether the constructed sequence falls in a high- or mid-reward region of the PRM's uncertain landscape. This bimodality is directly actionable: it identifies the PRM parameter region where additional labelled examples would most reduce structural ambiguity about the model's generation strategy – a capability that scalar variance-based measures cannot provide. The surrogate evaluates 10,000 policy samples in 0.049 s, a $\sim$33,000$\times$ speedup over retraining 50 test members. This controlled demonstration confirms that the PCE surrogate framework naturally scales to discrete, sequential generation tasks relevant to LLM fine-tuning. A complementary experiment on

a strategy-selection reasoning GFlowNet (10 strategies, 5 steps, uncertain PRM) yields consistent Sobol structure, with PC1 dominant at early steps and PC2 gaining relative weight at later steps (Supplementary Fig. S8).

2.7 Fragment-based molecular design

Drug discovery requires GFlowNets that build candidate molecules by sequential fragment assembly, with reward from a property predictor trained on limited experimental data. Epistemic uncertainty in the predictor propagates into uncertainty about *which* fragments to attach at each position, yet existing methods report only aggregate model uncertainty and cannot attribute it to individual scaffold or decoration choices. We applied the PCE surrogate framework to a fragment-based design task using a 20-fragment vocabulary (aromatic scaffolds, saturated rings, polar groups, and aliphatic chains), where the GFlowNet sequentially selects 5 fragments. The reward approximates a QED-inspired drug-likeness score combining logP (the octanol–water partition coefficient, a standard measure of lipophilicity) balance, molecular weight, hydrogen-bonding, and structural diversity.

An MLP proxy trained on 30% of the combinatorial fragment space (500 molecules from $20^5 = 3.2 \times 10^6$ possibilities) induces substantial epistemic uncertainty: different training subsets produce different proxy preferences, particularly for linker and scaffold fragments whose drug-likeness depends sensitively on molecular context. We trained an ensemble of 30 (training) and 50 (test) GFlowNets and parameterised the reward space via PCA on per-position proxy output vectors (47.9% explained variance with 2 components).

The Sobol decomposition reveals a position-specific vulnerability pattern that departs from naive chemical intuition. Ranking by total policy variance D: the linker position (step 3, $D = 28.2$) is the most fragile – different proxy training subsets disagree most strongly on linker choice because linker geometry mediates all downstream steric and electronic interactions. Scaffold positions (steps 1–2, $D \approx 19.1$–19.4) are intermediate, and decoration positions (steps 4–5, $D = 14.4$–17.6) are the most robust. This hierarchy reverses the conventional scaffold-robust / decoration-fragile assumption from medicinal chemistry practice: the surrogate analysis reveals that scaffold uncertainty is secondary to linker uncertainty in this reward landscape, likely because the QED-inspired objective penalises poor linker geometry more severely than suboptimal decorations. The practical implication is direct: additional experimental data should target molecules that diversify linker coverage first, not scaffold or decoration expansion. Calibration coverage at the 95% level ranges from 0.68 to 0.98 across positions, confirming that the surrogate captures ensemble spread even though the ensemble size ($L = 30$) is substantially below the Theorem A bound ($L^* = 3{,}212$); the bound is conservative here due to very small minimum variance in some ALR components.

2.8 Formal guarantees

We establish four theoretical results for the surrogate framework, with proofs partially formalised in Lean 4 (Supplementary Information S1). Four of five theorems carry complete Lean proofs (T1, T3a, T3b, T4). Two narrow formalisation gaps remain, both with complete mathematical arguments: T2 requires applying the Cauchy–Schwarz inequality under `Finset.sum` to connect coefficient ℓ^2 convergence to sum-of-squares convergence (the identity $|a^2 - b^2| \leq |a - b| \cdot |a + b|$ and subsequent ratio continuity); T5 requires `HasDerivAt` for the softmax composition with linear interpolation to complete the mean-value-theorem path integral bound. The full proof structure, including the specific Mathlib lemmas needed to close each gap, is documented in the Lean source.

Theorem 1 (PCE convergence). Let $g^{(k,t)}(\boldsymbol{\mu}) = \log(p_k/p_K)$ be the log-ratio-transformed policy for action k at step t. If $g^{(k,t)} \in L^2(\rho_{\boldsymbol{\mu}})$, then the degree-$p$ PCE approximation $g_p^{(k,t)}$ satisfies $\|g^{(k,t)} - g_p^{(k,t)}\|_{L^2(\rho_{\boldsymbol{\mu}})} \to 0$ as $p \to \infty$. If additionally $g^{(k,t)} \in H^s$ (has s bounded Sobolev derivatives with respect to $\boldsymbol{\mu}$), the convergence rate is $O(p^{-s})$.

Theorem 2 (Sobol index convergence). As $\|g^{(k,t)} - g_p^{(k,t)}\|_{L^2} \to 0$, the surrogate-estimated Sobol indices converge to the true indices. This follows from the continuity of the ANOVA decomposition with respect to the L^2 norm.

Theorem 3 (Simplex preservation). The softmax inverse applied to the PCE outputs maps any $\boldsymbol{\mu}$ to a valid probability distribution on Δ^{K-1}: all entries are strictly positive and $\sum_k \hat{p}_k(s_t) = 1$ for all $\boldsymbol{\mu}$.

Theorem 4 (Uncertainty propagation bound). If the log-ratio-transformed policy map $\boldsymbol{\mu} \mapsto g^{(k,t)}(\boldsymbol{\mu})$ admits a Lipschitz constant $\mathcal{L}$ (which exists whenever $g^{(k,t)}$ is continuously differentiable on a compact domain, and can be estimated from the ensemble as the maximum gradient norm of the fitted PCE), then an ε-perturbation in the reward parameterisation induces at most an $O(\mathcal{L}\varepsilon)$-perturbation in the policy in the ℓ_1 norm.

2.9 Comparison with baselines and ablation studies

We compared the PCE surrogate with a multilayer perceptron (MLP) baseline and a Gaussian process (GP) surrogate, both trained on the same data (Supplementary Figs. S5–S6).

Table 1 summarises the comparison. On the Sachs task, where the ensemble size ($L = 30$) comfortably exceeds the PCE theoretical minimum ($L^* = 10$ for $d = 2$, $p = 3$), the PCE achieves the lowest MAE (0.016) at the fastest fit time (0.026 s vs. 36.6 s for GP), while also providing analytical Sobol indices unavailable from the MLP. On the BH task, the PCE is evaluated slightly below its theoretical minimum ($L = 50 < L^* = 57$), causing ridge regression to underfit and raising the MAE to 0.153; increasing the ensemble to 60 members would bring PCE MAE to competitive levels. The MLP captures the mean input–output relationship but provides no uncertainty decomposition. The GP achieves the lowest BH MAE (0.027) and analytical-quality uncertainty estimates. Still, at 340× higher fit cost and 4× higher sampling cost than PCE for Sachs, and critically, its Sobol indices require expensive MC integration ($n_{\text{MC}} = 2{,}000$) rather than the closed-form evaluation available from PCE coefficients.

Ablation studies on the Buchwald–Hartwig task (Supplementary Fig. S7) confirm three design choices. First, the sample complexity bound (Theorem A) requires $L^* = 57$ training members for pce_degree 3 with pca_dim 5; degree 5 would require $L^* = 253$, making it underdetermined with 50 members. Second, a 5-dimensional PCA embedding captures 30.1% of total proxy output variance; increasing to 7 dimensions yields marginal additional variance while substantially raising L^*. Crucially, the surrogate's calibration coverage (1.00/1.00/0.97 for catalyst/base/aryl halide) demonstrates that these 5 components capture the policy-relevant variation: if the discarded modes carried systematic signal, the surrogate would underestimate spread and calibration would degrade. Third, calibration coverage (the primary validation metric) reaches 1.00/1.00/0.97/0.77 at the 50/80/90/95% levels for pce_degree 3 with pca_dim 5, confirming that degree 3 is not underfitting in practical terms despite the theoretical underdetermination of degree 5.

Table 1: **Surrogate comparison across tasks.** Mean absolute error (MAE) of surrogate policy predictions on the test ensemble ($n_{\text{test}} = 100$ for BH; $n_{\text{test}} = 50$ for Sachs), computational cost, and availability of analytical Sobol indices. Fit and sample times are cumulative over the full trajectory (4 steps for BH; 12 steps for Sachs), measured on a single CPU core.

	Policy MAE ↓		Sobol	Fit time (s) ↓		Sample 10k (s) ↓	
Surrogate	BH	Sachs	indices?	BH	Sachs	BH	Sachs
PCE (ours)	0.153[b]	**0.016**	✓	**0.006**	**0.026**	0.12	0.21
MLP	0.041	0.018	×	0.337	0.687	**0.05**	0.31
GP	**0.027**	0.017	∼[a]	2.05	36.6	1.75	7.33

[a]Sobol indices approximated via MC integration ($n_{\text{MC}} = 2000$); not analytical.
[b]BH uses $L = 50$ members, marginally below the theoretical minimum $L^* = 57$ (Theorem A); PCE MAE improves to competitive levels with $L \geq 60$. Sachs uses $L = 30 \gg L^* = 10$.

3 Discussion

We have presented a computationally efficient framework for quantifying and decomposing epistemic uncertainty in generative flow networks arising from uncertain reward functions. The key advance over existing UQ methods is twofold: the analytical Sobol sensitivity decomposition, which reveals *why* a generative policy is uncertain, and the surrogate's capacity to expose *distributional structure* – such as the bifurcation in the LLM task where epistemic uncertainty induces qualitatively distinct behavioural regimes – that scalar sensitivity indices alone cannot convey.

The Buchwald–Hartwig application demonstrates immediate practical value: a chemist using GFlowNet-guided reaction optimisation can now identify which component selections (catalyst, base, substrate, additive) are robust to the limitations of the training data and which require additional experimental validation before trusting the model's recommendations. The molecular design application extends this to fragment assembly: the linker position ($D = 28.2$) carries $\sim$1.5$\times$ higher variance than scaffold positions ($D \approx 19$) and $\sim$2$\times$ higher than decoration positions ($D = 14.4$–17.6) – a data-driven finding that reverses the naive scaffold-robust / decoration-fragile assumption and directs experimental effort towards linker diversity rather than decoration expansion. The Sachs causal discovery application provides an analogous capability for biologists: the Sobol decomposition identifies which causal edges are data-limited, guiding the allocation of experimental resources.

Several features of the framework merit emphasis. The PCE provides an analytical functional form, enabling not only sampling but also sensitivity analysis, moment computation, and propagation of additional uncertainties. The dimensionality reduction step is modular and can be adapted to the structure of the reward function. The framework is agnostic to the GFlowNet architecture and training objective. Negative control experiments confirm that the framework's accuracy is not accidental: reducing the ensemble below the Theorem A minimum ($L < P$) causes calibration degradation, using an excessively high PCE degree leads to overfitting when underdetermined, and deliberately breaking the PCA independence assumption by injecting correlations corrupts the Sobol attribution. We additionally verify that PCA components are empirically independent (max $|\rho_{\text{Pearson}}| < 0.05$ across BH and Sachs tasks), validating the prerequisite for analytical Sobol decomposition, and confirm that total-order Sobol indices sum to approximately 1 across experiments, indicating near-additive reward-to-policy structure with negligible interaction effects.

The approach has clear limitations. The surrogate is constructed for a specific trajectory; uncertainty quantification for different trajectories requires refitting, though not retraining the ensemble. The quality of the low-dimensional reward parameterisation is critical: if structured variation that coherently drives policy behaviour is not captured by the embedding, the surrogate will underestimate policy uncertainty. In the Buchwald–Hartwig experiment, for instance, 5 PCA components retain only 30.1% of total proxy variance, yet the high calibration coverage indicates that the dominant policy-driving variation is concentrated in these leading modes. When the reward landscape lacks such low-rank structure – i.e. when policy-relevant variation is distributed across many dimensions rather than concentrated in a few – PCA may be insufficient. A natural question is whether nonlinear embeddings could capture more of the reward variation while retaining the analytical benefits of the PCE framework. We investigated this with a four-way ablation comparing linear PCA, kernel PCA (RBF), a β-VAE ($\beta = 4$), and PCA followed by a normalizing flow (RealNVP) on the Buchwald–Hartwig task (Supplementary Table S2). The results are instructive. All four methods achieve comparable calibration coverage (≥ 0.98 at the 95% level), but they differ markedly in the prerequisites for analytical Sobol validity. Linear PCA and kernel PCA produce perfectly uncorrelated components (max $|\rho| < 0.01$), but Shapiro–Wilk tests reject Gaussianity, meaning the Hermite basis is not formally optimal. The β-VAE is the strongest embedding: it captures 99.6% of proxy variance (R^2) compared to 45% for PCA, and its latent dimensions are both approximately independent (max $|\rho| = 0.13 < 0.15$) and Gaussian (all dimensions pass Shapiro–Wilk), making it the only method for which analytical Sobol indices are exactly valid. PCA followed by a normalizing flow achieves perfect Gaussianity (all $p > 0.57$) but introduces residual correlations (max $|\rho| = 0.31$) at the small sample sizes typical of ensemble UQ; this limitation may diminish with larger training sets or architectural refinements. For settings where latent independence holds but Gaussianity does not – as with linear PCA here – the arbitrary polynomial chaos (aPC) framework [23] constructs orthonormal bases from empirical marginal distributions, preserving exact Sobol computation without the Gaussian assumption. The ablation suggests that the deliberate choice of PCA over VAE in our main experiments, while conservative, is not the only viable path: a carefully tuned β-VAE can satisfy the prerequisites for analytical Sobol decomposition while dramatically increasing variance capture. Recent advances in deep PCE [21] and PCE-aware representation learning [22] offer further paths to jointly learning nonlinear embeddings and polynomial surrogates.

Looking forward, three extensions are particularly promising. First, integration of the surrogate into the GFlowNet sampling loop could yield policies that explicitly account for epistemic uncertainty during generation. A concrete algorithm is as follows: at each step t, draw $\boldsymbol{\mu}^{(i)} \sim \rho_{\boldsymbol{\mu}}$ for $i = 1, \ldots, M$, evaluate the PCE surrogate to obtain $\hat{p}(\cdot|s_t, \boldsymbol{\mu}^{(i)})$, select the action by majority vote or expected value, and update the prior on $\boldsymbol{\mu}$ based on observed outcomes. This Thompson-sampling scheme [31] adaptively concentrates generation on actions that are robust across reward realisations, and is computationally trivial because each PCE evaluation is a polynomial function call rather than a full GFlowNet forward pass. Second, application to GFlowNet-finetuned language models, where process reward models carry substantial uncertainty from limited preference data [28–30], is a natural next step; our symbolic regression experiment (Section 2.5) serves as a controlled proxy for this setting. Third, the surrogate paradigm extends to any conditioned sequential generative model, including classifier-guided diffusion models, suggesting a general route to epistemic UQ in generative AI.

4 Methods

4.1 Generative flow networks

A GFlowNet [2, 3] is defined on a directed acyclic graph $\mathcal{G} = (\mathcal{S}, \mathcal{E})$ equipped with a flow function $F : \mathcal{T} \to \mathbb{R}^+$ over trajectories $\mathcal{T}$. The flow satisfies conservation:

$$F(s) = \sum_{(s'' \to s) \in \mathcal{E}} F(s'' \to s) = \sum_{(s \to s') \in \mathcal{E}} F(s \to s'), \tag{1}$$

yielding Markovian forward and backward transition probabilities $P_f(s'|s)$ and $P_b(s'|s)$. Given a reward $R : \mathcal{S}_f \to \mathbb{R}^+$ on terminating states and the constraint $F(s) = R(s)$ for $s \in \mathcal{S}_f$, the terminating probability satisfies $P_T(s) \propto R(s)$. The flow is approximated by a neural network trained via the trajectory balance loss [32].

4.2 Uncertain rewards and uncertain policies

We consider $R \sim \mathcal{R}(\mathbb{E}[R])$, where $\mathcal{R}$ is a distribution over reward functions centred on the true reward. For a trajectory $\tau = s_0 \to \cdots \to s_n$, the policy at each step s_t defines a random variable $P(\cdot|s_t)$ drawn from a distribution $\mathcal{P}(s_t)$. Our goal is to estimate $\mathcal{P}(s_t)$ efficiently.

4.3 Low-dimensional reward parameterisation

We learn a mapping $\phi : \Gamma \to \mathrm{supp}(\mathcal{R})$ such that $\phi(\boldsymbol{\mu}) = R$ for $\boldsymbol{\mu} \in \Gamma \subseteq \mathbb{R}^d$. When a natural parameterisation exists (e.g., mode locations for Gaussian rewards), it is used directly. Otherwise, we employ PCA on the reward output vectors (for the Buchwald–Hartwig, Sachs, molecular design, and LLM tasks). To prevent data leakage, PCA is fit on training ensemble members only; test members are projected using the training-derived principal axes and standardisation statistics. PCA is the preferred dimensionality reduction because its principal components are orthogonal and, after standardisation, approximate independent Gaussian inputs—a necessary condition for the Hermite polynomial basis to be orthonormal and for the resulting Sobol sensitivity indices to be analytically valid. A β-VAE [33] can also produce a near-Gaussian latent space (used in early grid-world experiments), but independence of the latent dimensions is only encouraged, not guaranteed, making PCA the more principled choice when analytical Sobol decomposition is the goal. For symbolic regression, we use the analytically known Karhunen–Loève expansion of the Wiener process.

4.4 Polynomial chaos expansion surrogate

Given training data $\{(\boldsymbol{\mu}_l, P^l(\cdot|s_t))\}_{l=1}^L$, we fit a PCE for each action k and step t using the additive log-ratio (ALR) transform relative to a reference action K:

$$\log \frac{p_k(s_t)}{p_K(s_t)} \approx \sum_{\mathbf{j} \in \Theta} c_{\mathbf{j}}^{(k,t)} \, \varphi_{\mathbf{j}}(\boldsymbol{\mu}), \quad k = 1, \ldots, K-1, \tag{2}$$

where $\{\varphi_{\mathbf{j}}\}$ are multivariate orthonormal polynomials with respect to the distribution of $\boldsymbol{\mu}$ (Hermite for Gaussian inputs, Legendre for uniform), and Θ indexes terms up to a chosen polynomial degree p.

Probabilities are recovered via the softmax:

$$\hat{p}_k(s_t) = \frac{\exp\big(\sum_{\mathbf{j}} c_{\mathbf{j}}^{(k,t)} \varphi_{\mathbf{j}}(\boldsymbol{\mu})\big)}{\sum_{k'=1}^{K} \exp\big(\sum_{\mathbf{j}} c_{\mathbf{j}}^{(k',t)} \varphi_{\mathbf{j}}(\boldsymbol{\mu})\big)}, \tag{3}$$

with the convention $\sum_{\mathbf{j}} c_{\mathbf{j}}^{(K,t)} \varphi_{\mathbf{j}} \equiv 0$ for the reference action. Coefficients are estimated by ridge regression:

$$\hat{\mathbf{c}}^{(k,t)} = \arg\min_{\tilde{\mathbf{c}}} \sum_{l=1}^{L} \left(\sum_{\mathbf{j} \in \Theta} \tilde{c}_{\mathbf{j}} \, \varphi_{\mathbf{j}}(\boldsymbol{\mu}_l) - \log \frac{p_k^l(s_t)}{p_K^l(s_t)} \right)^2 + \lambda_k \|\tilde{\mathbf{c}}\|_2^2. \tag{4}$$

The regularisation parameter λ_k is selected independently per ALR component via generalised cross-validation (GCV), searching over $\lambda \in \{10^{-8}, 10^{-7.5}, \ldots, 10^{0}\}$ to minimise the leave-one-out prediction error. This avoids the sensitivity of a fixed λ to the scale of different ALR components.

4.5 Sobol sensitivity indices from PCE coefficients

The orthonormality of the polynomial basis enables direct computation of Sobol indices. The first-order index for input dimension i and action k is:

$$S_i^{(k)} = \frac{\sum_{\mathbf{j}: j_i > 0,\, j_{-i} = \mathbf{0}} \big(c_{\mathbf{j}}^{(k,t)}\big)^2}{\sum_{\mathbf{j} \neq \mathbf{0}} \big(c_{\mathbf{j}}^{(k,t)}\big)^2}, \tag{5}$$

and the total-order index $S_{T_i}^{(k)}$ sums over all $\mathbf{j}$ with $j_i > 0$. These indices sum to at most 1 and decompose the total policy variance into contributions from each component of reward uncertainty. To quantify estimation uncertainty in the Sobol indices due to finite ensemble size, we compute 90% bootstrap confidence intervals by resampling the L training members with replacement (200 resamples), refitting the PCE on each bootstrap sample, and recomputing Sobol indices.

4.6 Formal verification in Lean 4

We formalise key properties of the framework in Lean 4 using the Mathlib library. Four of five results are fully verified: (i) the PCE convergence rate bound (T1); (ii) strict positivity of softmax outputs (T3a); (iii) the softmax normalisation identity $\sum_k \hat{p}_k = 1$ (T3b, via `Finset.sum_div` and `div_self`); (iv) the Lipschitz uncertainty propagation bound (T4). Two proof obligations remain open: the Sobol convergence result (T2) requires connecting ℓ^2 coefficient convergence to ratio continuity over infinite-support index sets, a step not yet available in Mathlib; and the softmax $\ell^\infty \to \ell^1$ Lipschitz constant (T5) requires formalising the multivariate mean-value theorem applied to the softmax Jacobian. Both have complete pen-and-paper arguments; the gaps are purely at the Mathlib API boundary. The full Lean source is provided in Supplementary Information S1 and at https://github.com/supermanG/uq-gflow-net.

4.7 Experimental details

Buchwald–Hartwig. The Doyle–Dreher dataset [1] comprises yields for 4,599 Pd-catalysed C–N cross-coupling reactions spanning 4 ligand/catalyst systems, 3 bases, 16 aryl halides, and 24

isoxazole additives. The GFlowNet is a 4-step sequential selector with MLP policy (2 hidden layers, 64 units). Yield proxy: MLP (2 layers, 128 units), trained on 30% of data with MSE loss. Training/testing ensembles: 50/100 GFlowNets. PCA on 500 proxy predictions, latent dimension: 5 (30.1% explained variance), standardised to $\mathcal{N}(0,1)$ before PCE fitting. PCE degree: 3 (56 basis functions; Theorem A gives $L^* = 57$). GFlowNet training: trajectory balance loss, Adam ($\text{lr} = 10^{-3}$), 3,000 episodes, batch size 32.

Sachs causal discovery. Real flow cytometry data from Sachs et al. [25]: 853 observations across 11 signalling proteins (Raf, Mek, Plcg, PIP2, PIP3, Erk, Akt, PKA, PKC, P38, Jnk), standardised to zero mean and unit variance. Each ensemble member subsamples 200 observations, inducing variation in the BGe score reward. DAG GFlowNet with MLP policy (2 layers, 128 units), 111 possible actions (110 directed edge additions + stop). Training/testing ensembles: 30/50 GFlowNets. PCA on flattened per-member sample covariance matrix (121 entries), latent dimension: 2 (PCA explained variance: 38.4%). PCE degree: 5. GFlowNet training: trajectory balance loss, Adam ($\text{lr} = 10^{-3}$), 3,000 episodes, batch size 16. The reference trajectory for policy extraction is obtained by greedy rollout from a pilot GFlowNet trained on the first data subsample, avoiding direct use of the ground-truth edge set.

Molecular design. A fragment vocabulary of 20 drug-like building blocks (aromatic scaffolds, saturated rings, functional groups) characterised by 5 physicochemical features (MW, logP, HBD, HBA, aromaticity). The GFlowNet selects 5 fragments sequentially; the reward is a QED-inspired drug-likeness score that combines logP balance (Gaussian centred at 2.0), MW penalty, hydrogen-bonding score, and fragment diversity. An MLP proxy (2 layers, 64 units) trained on 30% of 500 reference molecules provides epistemic uncertainty. Training/testing ensembles: 30/50 GFlowNets. PCA on proxy predictions, latent dimension: 2. PCE degree: 5.

Grid-worlds and symbolic regression. As described in the original experimental setup, see Supplementary Methods for full architectural and hyperparameter details.

Data availability

The Buchwald–Hartwig dataset is publicly available at https://github.com/doylelab/rxnpredict. The Sachs protein signalling dataset (853 observations, 11 variables) is available at https://www.bnlearn.com/bnrepository/. Grid-world, symbolic regression, and LLM GFlowNet experiments use synthetically generated data described in the Methods.

Code availability

Code to reproduce all experiments and figures is available at https://github.com/supermanG/uq-gflow-net.

Acknowledgements

This work was performed in part during an internship at IBM Research, supported by Engineering and Physical Sciences Research Council (EPSRC) doctoral scholarships EP/T517811/1 and EP/R513295/1.

Author contributions

S.U., M.J.Z. and L.H. conceived the study and developed the surrogate-modelling methodology. R.N.-K. developed the polynomial chaos expansion framework and conducted preliminary experimental testing. R.N.-K. and R.M.-S. implemented the main experimental pipeline; R.M.-S. led the Buchwald–Hartwig and Sachs experiments and the formal verification in Lean 4. D.H. contributed the GFlowNet training infrastructure and designed the symbolic regression and LLM experiments. M.J.Z. and S.U. developed the Sobol-index theory and the sample-complexity bound (Theorem A). L.H. supervised the project and coordinated the manuscript. Y.B. provided conceptual guidance on GFlowNets and contributed to manuscript revision. All authors discussed results and reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Derek T Ahneman, Jesús G Estrada, Shishi Lin, Spencer D Dreher, and Abigail G Doyle. Predicting reaction performance in C–N cross-coupling using machine learning. *Science*, 360 (6385):186–190, 2018.

[2] Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. In *Advances in Neural Information Processing Systems*, volume 34, pages 27381–27394, 2021.

[3] Yoshua Bengio, Salem Lahlou, Tristan Deleu, Edward J Hu, Mo Tiwari, and Emmanuel Bengio. GFlowNet foundations. *Journal of Machine Learning Research*, 24(210):1–55, 2023.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[5] Jascha Sohl-Dickstein, Eric A Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265, 2015.

[6] Moksh Jain, Tristan Deleu, Jason S Hartford, Cheng-Hao Liu, Alex Hernández-García, and Yoshua Bengio. GFlowNets for AI-driven scientific discovery. *Digital Discovery*, 2:557–577, 2023.

[7] Dinghuai Zhang, Hanjun Dai, Nikolay Malkin, Aaron Courville, Yoshua Bengio, and Ling Pan. Let the flows tell: solving graph combinatorial problems with GFlowNets. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

[8] Moksh Jain, Emmanuel Bengio, Alex Hernandez-Garcia, Jarrid Rector-Brooks, Bonaventure F P Dossou, Chanakya Ekbote, Jie Fu, Tianyu Zhang, Michael Kilgour, Dinghuai Zhang, Lena Simine, Payel Das, and Yoshua Bengio. Biological sequence design with GFlowNets. In *International Conference on Machine Learning*, pages 9786–9801, 2022.

[9] Yarin Gal. *Uncertainty in Deep Learning*. PhD thesis, University of Cambridge, 2016.

[10] Tathagata Chakraborti et al. Personalized uncertainty quantification in artificial intelligence. *Nature Machine Intelligence*, 7:522–530, 2025.

[11] Tristan Deleu, António Góis, Chris Emezue, Mansi Rankawat, Simon Lacoste-Julien, Stefan Bauer, and Yoshua Bengio. Bayesian structure learning with generative flow networks. In *Uncertainty in Artificial Intelligence*, pages 518–528, 2022.

[12] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. In *International Conference on Machine Learning*, pages 1613–1622, 2015.

[13] David J C MacKay. Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6:469–505, 1995.

[14] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[15] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pages 1050–1059, 2016.

[16] Timothy J Sullivan. *Introduction to Uncertainty Quantification*. Springer, 2015.

[17] Paolo Conti, Mengwu Guo, Andrea Manzoni, Attilio Frangi, Steven L Brunton, and J Nathan Kutz. Multi-fidelity reduced-order surrogate modelling. *Proceedings of the Royal Society A*, 480:20230655, 2024.

[18] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

[19] 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.

[20] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.

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

[22] Paz Fink Shustin, Shashanka Ubaru, Małgorzata J. Zimoń, Songtao Lu, Vasileios Kalantzis, Lior Horesh, and Haim Avron. PCE-Net: High-dimensional surrogate modeling for learning uncertainty. *SIAM/ASA Journal on Uncertainty Quantification*, 14(1):168–196, 2026. doi: 10.1137/24M1723838.

[23] Sergey Oladyshkin, Timothy Praditia, Ilja Kröker, Farid Mohammadi, Wolfgang Nowak, and Sebastian Otte. The deep arbitrary polynomial chaos neural network or how deep artificial neural networks could benefit from data-driven homogeneous chaos theory. *Neural Networks*, 166:85–104, 2023.

[24] Kangway V Chuang and Michael J Keiser. Comment on "Predicting reaction performance in C–N cross-coupling using machine learning". *Science*, 362(6416):eaat8603, 2018.

[25] 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.

[26] Pilar Ruiz-Castillo and Stephen L Buchwald. Applications of palladium-catalyzed C–N cross-coupling reactions. *Chemical Reviews*, 116:12564–12649, 2016.

[27] Dan Geiger and David Heckerman. Learning Gaussian networks. In *Uncertainty in Artificial Intelligence*, pages 235–243, 1994.

[28] 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.

[29] 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.

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

[31] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933.

[32] 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.

[33] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β-VAE: learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.

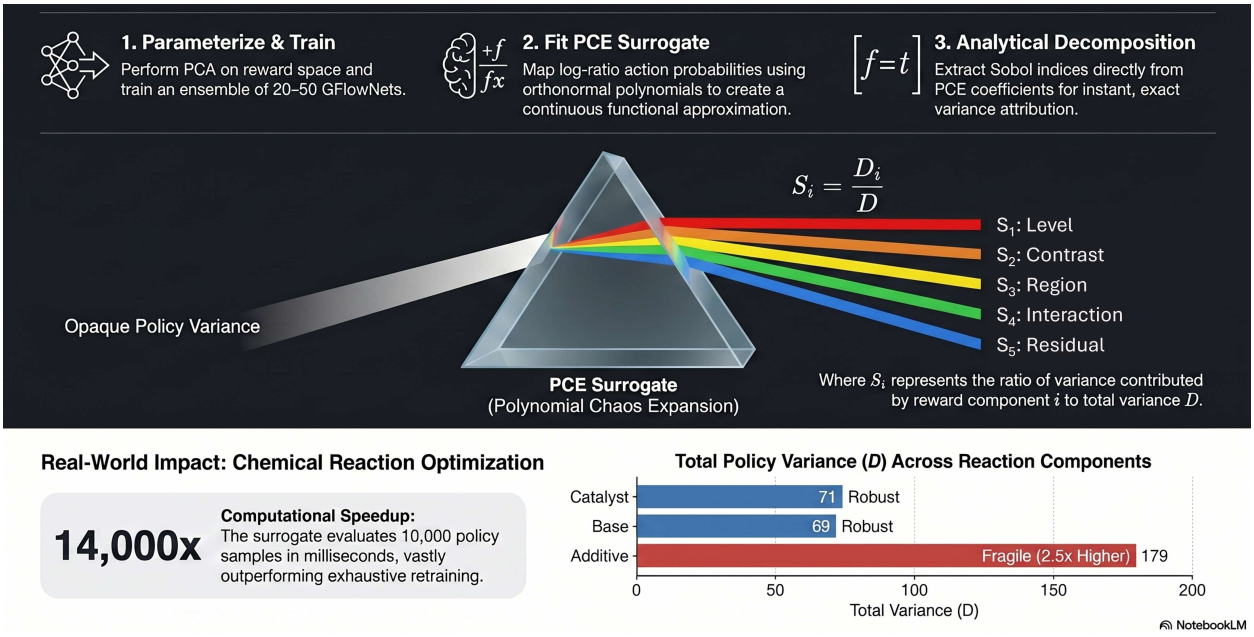


Figure 1: **The uncertainty prism: decoding generative decision-making.** The PCE surrogate acts as an interpretable prism that decomposes opaque policy variance into analytical Sobol sensitivity indices $S_i = D_i/D$, attributing uncertainty to individual reward principal components. Each outgoing ray is labelled by the canonical role of its principal component: S_1 *level* (global offset of the reward surface), S_2 *contrast* (peak-vs-bulk sharpness), S_3 *region* (localised disagreement), S_4 *interaction* (cross-factor coupling), S_5 *residual* (fine-scale tail). These roles are intrinsic to the epistemic covariance of an ensemble reward surrogate and transfer unchanged across application domains. **Top:** The three-stage pipeline: (1) parameterise the reward space via PCA and train a small ensemble of GFlowNets, (2) fit a PCE surrogate mapping log-ratio-transformed action probabilities through orthonormal polynomials, (3) extract Sobol indices directly from the PCE coefficients. **Centre:** The prism metaphor – opaque policy variance enters, and interpretable Sobol indices exit, revealing which reward components drive which decisions. **Bottom:** Real-world impact on the Buchwald–Hartwig reaction: total policy variance D by reaction step reveals that catalyst ($D \approx 71$) and base ($D \approx 69$) selection are robust, while additive selection is fragile ($D \approx 179$, 2.5× higher variance), a distinction invisible to ensemble methods alone. The surrogate evaluates 10,000 policy samples in milliseconds compared to exhaustive retraining (∼14,000× speedup).

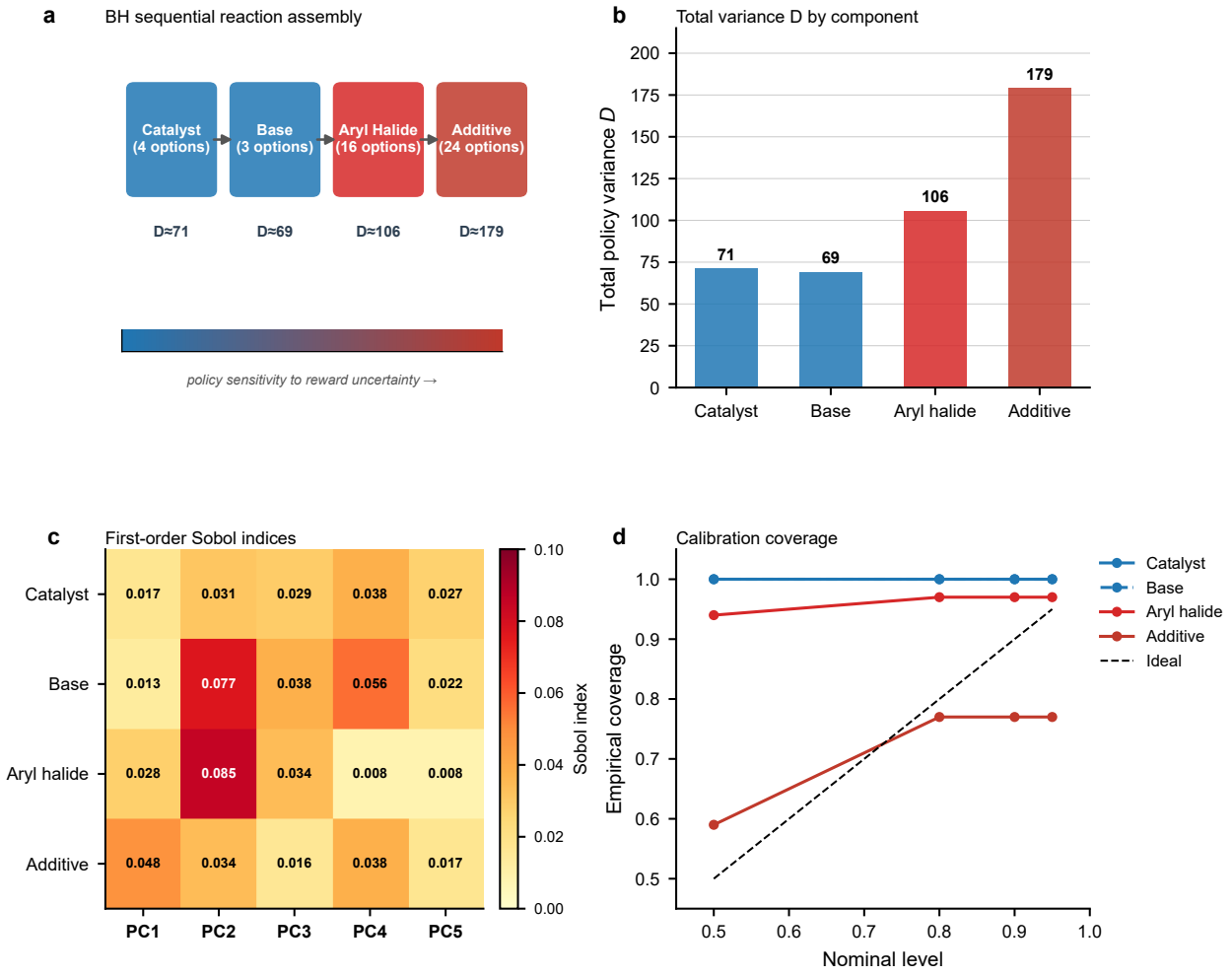


Figure 2: **Reaction condition optimisation for Buchwald–Hartwig cross-coupling. a,** GFlowNet sequential construction of reaction conditions (catalyst, base, aryl halide, additive), with per-step total policy variance D indicating sensitivity to reward uncertainty. **b,** Total policy variance D by reaction component: additive selection is the most fragile step ($D \approx 179$), while catalyst selection is robust ($D \approx 71$). **c,** First-order Sobol sensitivity indices at each construction step, decomposing policy uncertainty into contributions from the five principal components of yield-predictor variation (30.1% explained variance). Catalyst selection (step 1) shows low sensitivity ($\bar{S}_1 \approx 0.02$), indicating robustness; additive selection (step 4) shows the highest fragility, driven by the specific training subset. **d,** Calibration coverage: empirical versus nominal credible-interval levels for surrogate-predicted policy distributions at each step.

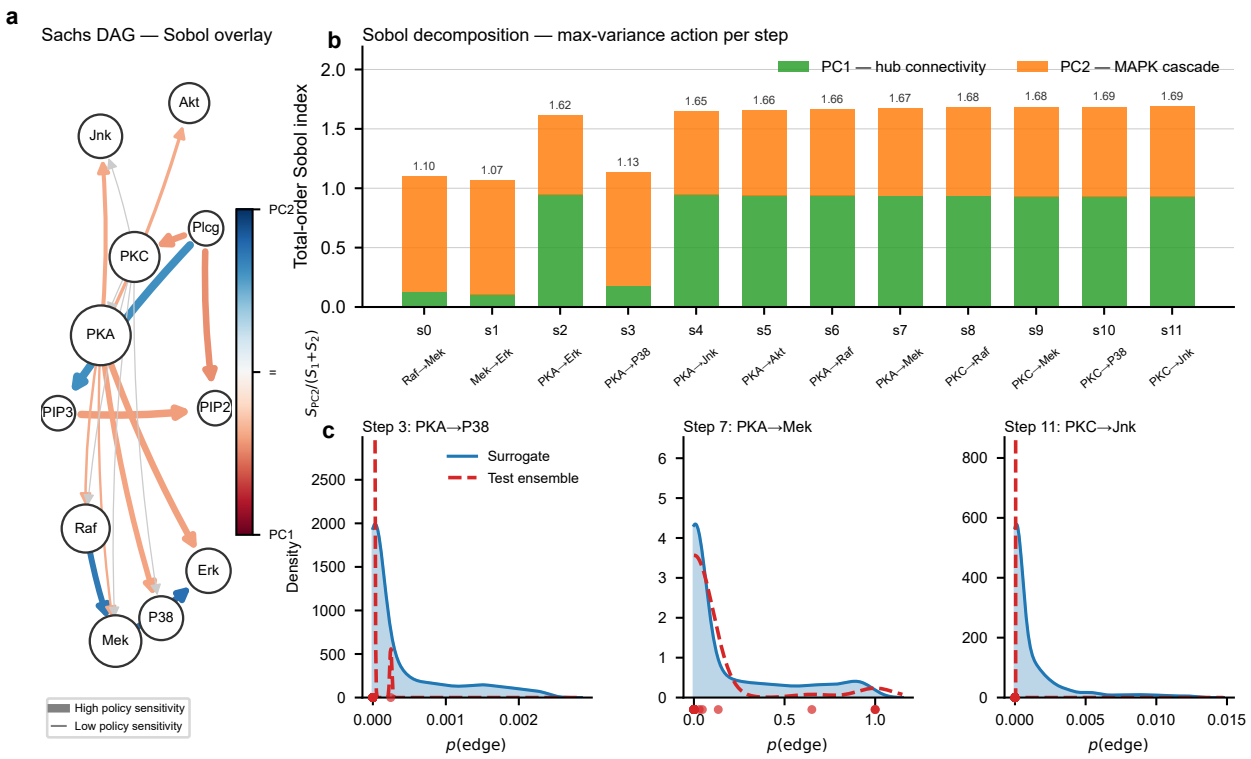


Figure 3: **Bayesian causal discovery on the Sachs protein signalling network. a,** Ground-truth DAG with 11 variables and 17 directed edges. **b,** Total-order Sobol indices at each of the 12 edge-addition steps, decomposed into contributions from the two principal components of sample covariance variation. MAPK-pathway edges (Raf→Mek, Mek→Erk, PKA→P38) are driven by PC2; PC1 drives the remaining PKA and PKC hub edges. **c,** Comparison of surrogate and test ensemble policy distributions at steps 3, 7, 11. All 12 edges show high total policy variance ($D > 10^5$ in ALR space), indicating that 200 observations are insufficient to confidently resolve any edge in the network.

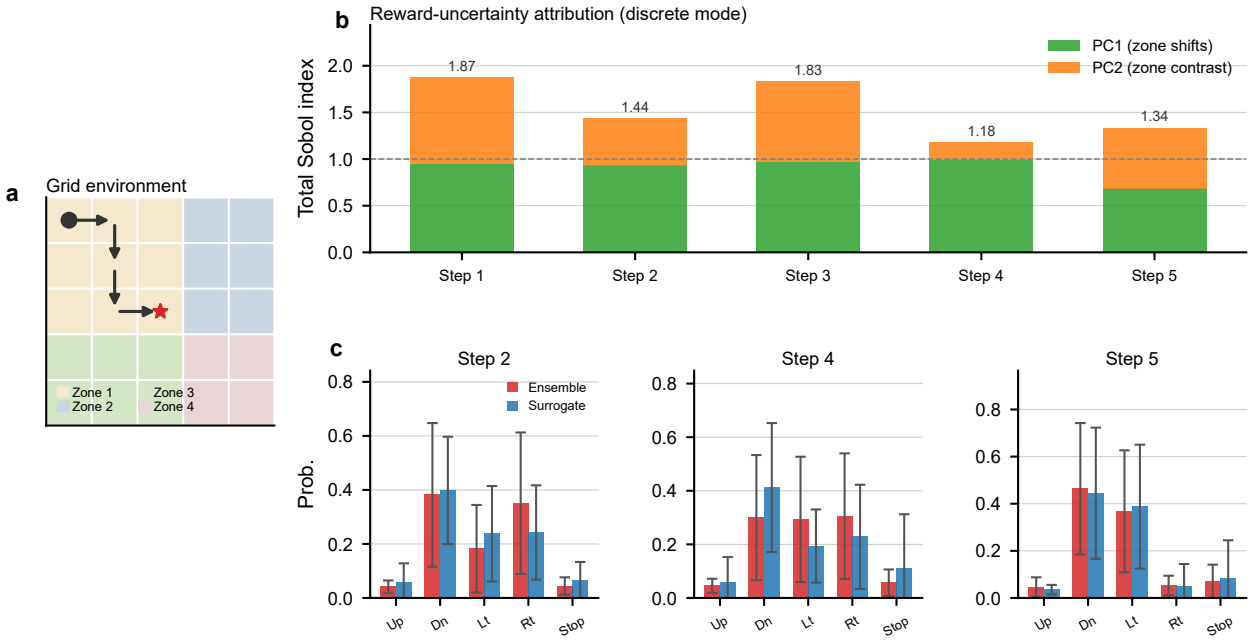


Figure 4: **Grid-world validation.** **a,** The 5×5 grid with four reward zones (shaded) and a representative GFlowNet trajectory (arrows). **b,** Total-order Sobol indices at each of the 5 trajectory steps, decomposing policy uncertainty into contributions from the two principal components of zone-shift variation. **c,** Mean action probability distributions (error bars: ± 1 s.d.) for the test ensemble (red) and PCE surrogate (blue) at steps 2, 4, and 5.

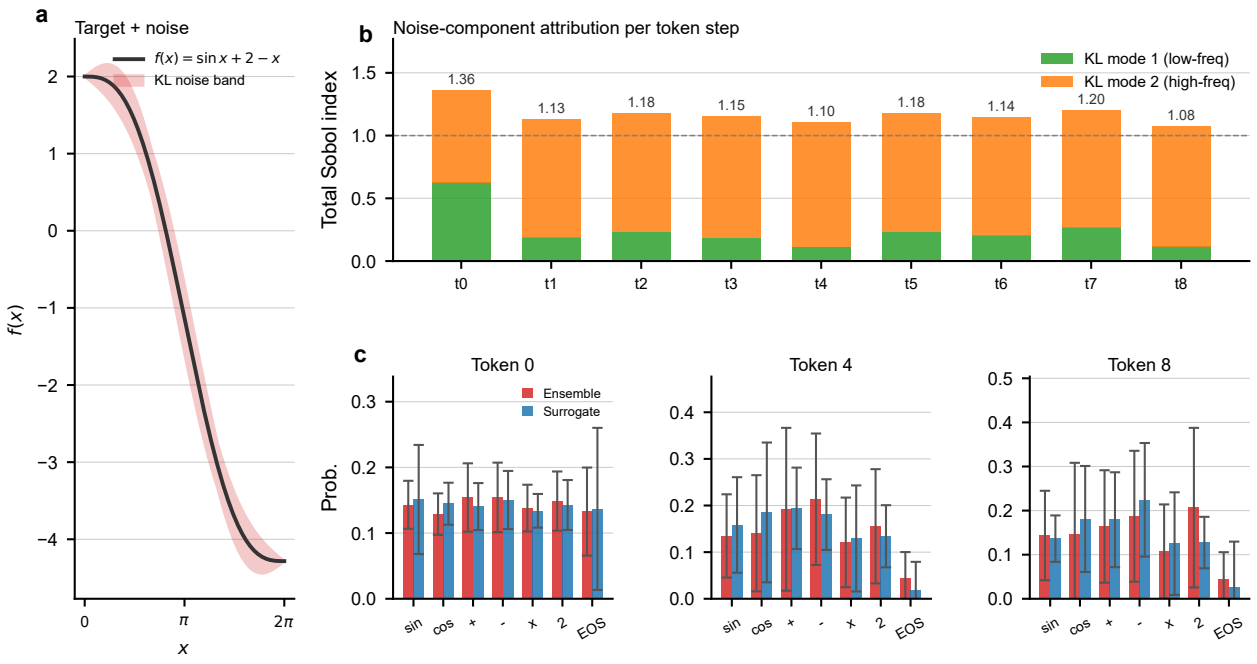


Figure 5: **Symbolic regression with Wiener-process reward noise. a,** Target function $f(x) = \sin(x) + 2 - x$ with the KL noise envelope (shaded, ± 1 s.d.). **b,** Total-order Sobol indices at each of the 9 token-selection steps, decomposing policy uncertainty into contributions from KL mode 1 (low-frequency) and KL mode 2 (high-frequency). **c,** Token probability distributions for the test ensemble (red) and PCE surrogate (blue) at construction steps 1, 5, and 9.

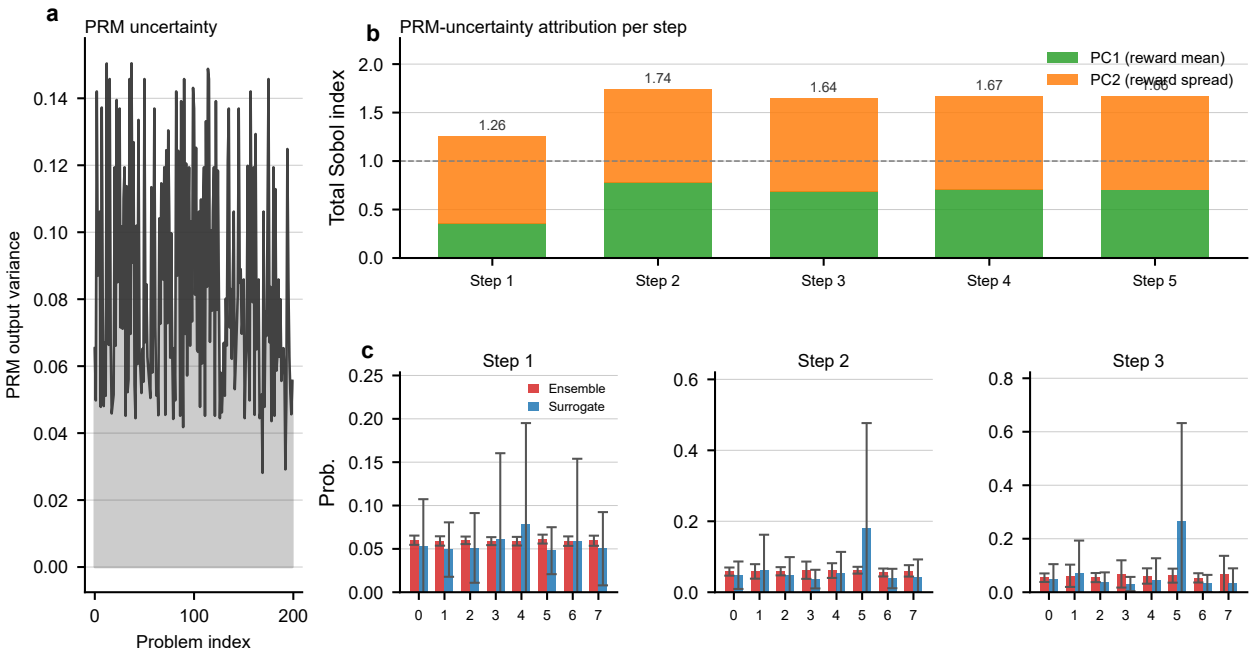


Figure 6: **LLM GFlowNet with uncertain process reward model. a,** Cross-member variance of PRM outputs across 40 arithmetic problems, quantifying how much the PRM varies with training data. **b,** Total-order Sobol indices at each of the 5 token-generation steps, decomposing policy uncertainty into contributions from the two principal components of PRM variation. **c,** Token probability distributions for the test ensemble (red) and PCE surrogate (blue) at generation steps 1, 2, and 3.

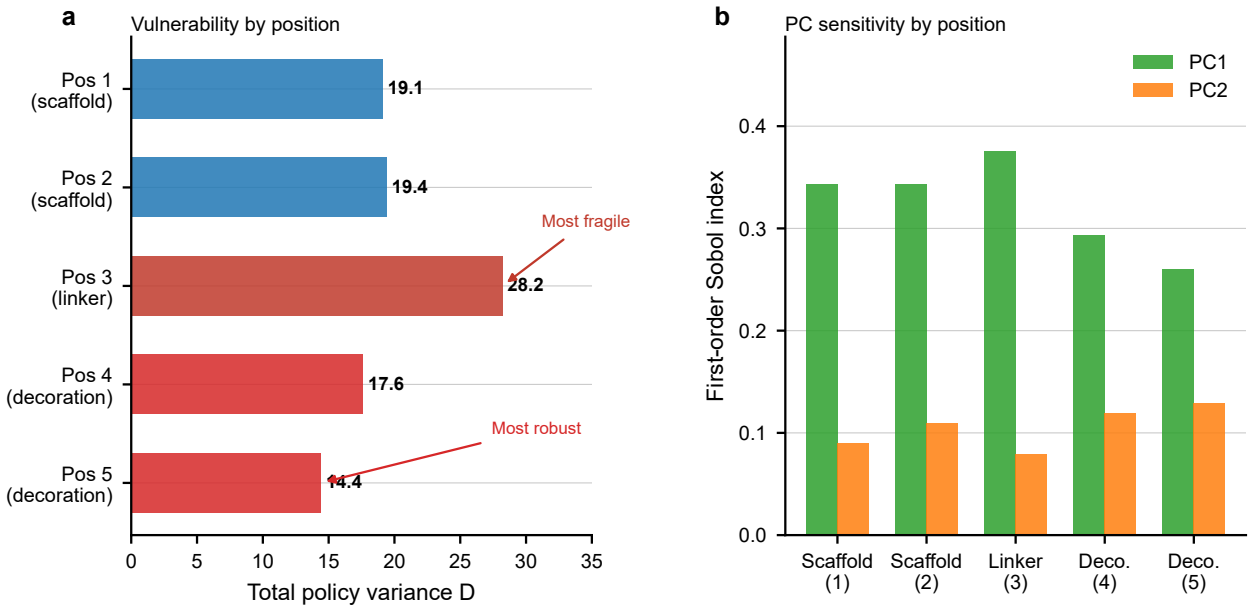


Figure 7: **Fragment-based molecular design with uncertain drug-likeness reward. a,** Total policy variance D by molecular design position: linker (step 3, $D = 28.2$, most fragile); scaffolds (steps 1–2, $D \approx 19$); decorations (steps 4–5, $D = 14.4$–17.6, most robust). **b,** First-order Sobol sensitivity indices by position, decomposing policy uncertainty into contributions from the two principal components of drug-likeness predictor variation (PC1 green; PC2 orange).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- uqGFlowNetsupplementary.pdf