Ծավալը՝ 20 էջ: Տպաքանակը՝ 70:
ՀՀ ԳԱԱ ԻԱՊԻ կոմպյուտերային պոլիգրաֆիայի լաբորատորիա:
Երևան, Պ. Սևակի 1

ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏ

Էդվարդ Արսենի Խալաֆյան

Էնտրոպիայով առաջնորդվող ԱԲ. հավանականային եզրահանգում, պատճառահետևանքային ներկայացումներ և մոդելների ադապտիվ ճշգրտում

Ե.13.05 – «Մաթեմատիկական մոդելավորում, թվային մեթոդներ և ծրագրերի համալիրներ» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության

ՍԵՂՄԱԳԻՐ

ԵՐԵՎԱՆ 2026

INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF THE NAS RA

Edvard Arsen Khalafyan

Entropy-Driven AI: Probabilistic Inference, Causal Representations, and Adaptive Model Fine-Tuning

SYNOPSIS

of the dissertation for obtaining a Ph.D. degree of Technical Sciences on specialty 05.13.05 "Mathematical modeling, digital methods and program complexes"

YEREVAN 2026

Ատենախոսության թեման հաստատվել է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում:

Գիտական ղեկավար՝ Ֆիզ. մաթ. գիտ. թեկնածու Ա. Ն. Հարությունյան
Պաշտոնական ընդդիմախոսներ՝ Ֆիզ. մաթ. գիտ. դոկտոր Է. Մ. Պողոսյան
Ֆիզ. մաթ. գիտ. թեկնածու Ա. Վ. Պողոսյան

Առաջատար կազմակերպություն՝
Ա. Ի. Ալիխանյանի անվան ազգային գիտական լաբորատորիա (Երևանի ֆիզիկայի ինստիտուտ)

Ատենախոսության պաշտպանությունը կայանալու է 2026 թ. մարտի 17-ին՝ ժամը 14:00-ին, ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտում գործող 037 մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ Երևան, 0014, Պ. Սևակի 1:

Ատենախոսությանը կարելի է ծանոթանալ ՀՀ ԳԱԱ ԻԱՊԻ գրադարանում:
Սեղմագիրն առաքված է 2026թ. փետրվարի 13-ին:

Մասնագիտական խորհրդի գիտական քարտուղար Ֆիզ. մաթ. գիտ. դոկտոր՝ Մ. Ե. Հարությունյան

The subject of the dissertation is approved by Institute for Informatics and Automation Problems of NAS RA

Scientific supervisor: Candidate of phys-math sciences A. N. Harutyunyan
Official opponents: Doctor of phys-math sciences E. M. Poghosyan
Candidate of phys-math sciences A. V. Poghosyan

Leading organization:
A. I. Alikhanyan national science laboratory (Yerevan Physics Institute)

The defense will take place on the 17th of March 2026 at 14:00, at the Specialized Council 037 «Informatics» at the Institute of Informatics and Automation Problems of NAS RA. Address: Yerevan, 0014, P. Sevak 1.

The dissertation is available in the library of IIAP NAS RA.
The synopsis is sent out on the 13th of February 2026.

Scientific secretary of the special counsel:
Ph.D.Ph.M.S. M. E. Haroutunian

General description of the work

Entropy is a foundational quantity across information theory, statistics, physics, and artificial intelligence. In information theory, Shannon entropy characterizes average code length and limits of lossless compression. In statistics, entropy is frequently used as an inference principle to reconstruct distributions from partial information such as empirical moments. In classical machine learning, decision tree algorithms often use the Shannon entropy (to compute information gain) to choose the best split. Similarly, entropy-based penalties have been proposed for model regularization, effectively discouraging overly confident (low-entropy) predictions. In modern large language models (LLMs), entropy is computed from the next-token predictive probability distribution and is used implicitly or explicitly for confidence estimation and the regulation of sampling and decoding strategies.

We argue that entropy becomes genuinely informative only when it is tied to a clearly defined probabilistic object and a clearly stated decision purpose. In many applied papers and engineering systems, entropy is introduced as a general synonym for uncertainty. That habit is risky. A single numerical entropy value can refer to different objects that live on different sample spaces. It can be computed before or after decoding transforms. It can be used for different decisions, such as estimation, selection, routing, and compute control. When these choices are not spelled out, entropy claims can look universal even though they are not.

To avoid that failure mode, we develop a simple and strict methodology. Every entropy-based argument is framed through three requirements. First, the probabilistic object must be specified. This includes the sample space and the distribution that lives on it. Second, the purpose must be specified. This means a decision problem with an explicit loss, utility, or measurable objective. Third, validation must be specified. Validation means that the claim is tested against outcomes and compared to a baseline that corresponds to not injecting structure that is unsupported by the data.

This object–purpose–validation discipline is applied across three domains. The first domain is finite-sample inference for categorical distributions through maximum-entropy (MaxEnt) methods. The second domain is interpretable representation learning, where nonnegative matrix factorization (NMF) is reinterpreted through a probabilistic and common-cause lens. The third domain is reliability and compute-aware adaptation in LLMs, where token entropy is examined as an uncertainty signal, then transformed into training and inference policies that allocate compute.

Throughout the work, entropy is coupled to Kullback–Leibler (KL) divergence because KL provides a direct bridge to decision risk and approximation error. Concretely, for a discrete distribution p on $\mathcal{X}$, Shannon entropy is

$$S(p) = -\sum_{x\in\mathcal{X}} p(x)\log p(x),$$

and KL divergence between q and p is

$$\mathrm{KL}(q\|p) = \sum_{x\in\mathcal{X}} q(x)\log\frac{q(x)}{p(x)}.$$

Entropy measures dispersion within one distribution. KL measures mismatch between two

istributions. Many practical decisions in inference, representation learning, and model evaluation an be written as minimizing expected KL or a closely related functional. This is why we use orward KL risk in the inference chapter. This is also why later chapters treat entropy as a candidate redictor of error risk and a candidate driver of compute allocation.

A key technical issue for the LLM chapters is the precise definition of the entropy object. Let $_{\le t}$ denote the input prompt (the sequence of tokens up to position t), and let $\mathcal{V}$ be the model's ull vocabulary of possible next tokens. The model assigns a probability $p_\theta(v \mid x_{\le t})$ to each token $\in \mathcal{V}$. The next-token entropy is then defined as

$$S_t = -\sum_{v \in \mathcal{V}} p_\theta(v \mid x_{\le t}) \log p_\theta(v \mid x_{\le t}),$$

where log is the natural logarithm. In other words, S_t is the Shannon entropy of the full output distribution over the vocabulary. (Any decoding-time operations like temperature scaling or truncation would change this distribution, so comparisons of S_t require fixing these choices.) By default, we se the entropy of the original, full-vocabulary distribution unless explicitly stated otherwise.

The broader motivation is contemporary. Language models are deployed in settings where eliability matters. Compute is also a hard constraint both at training time and at inference time. Entropy is one of the few signals that is cheap to compute and available inside most probabilistic nodels. We show when this signal is useful, when it is not, and how to convert validated entropy neasurements into decisions.

Relevance of the research topic

The work is relevant because it targets a recurring gap between conceptual language and operational practice. Entropy is routinely described as a universal measure of uncertainty. In reality, ncertainty is task-dependent and decision-dependent. An uncertainty signal is valuable only if it redicts something that matters, such as error probability, future loss, or the marginal benefit of xtra compute.

In finite-sample inference, sparse data regimes are common. In such regimes, it is easy to verfit noise by imposing constraints that look meaningful but are statistically unstable. Entropy-naximizing inference is often presented as conservative. We show that it is not automatically conervative. It can add structure that harms risk when constraints are noisy or misaligned. This natters because MaxEnt is still used in scientific modeling and in ML pipelines where constraints re estimated from limited data.

In representation learning, interpretability claims often collapse under perturbations. If basis elments change drastically across random seeds or small noise, then the representation is not reliable, ven if reconstruction error is small. By tying NMF to a probabilistic common-cause view and by ntroducing a predictability-based effective rank, we contribute tools to select representations that re stable and interpretable in practice.

In language model deployment, entropy is widely used as a confidence proxy. However, confience proxies must be validated across question types and failure modes. A single global correlaon is not enough. This work's stratified evaluations matter because they clarify when entropy can upport routing and abstention policies and when additional reasoning-aware signals are required.

Finally, compute-aware adaptation is an urgent theme. Distillation, reasoning supervision, and

multi-model pipelines all introduce explicit costs. We use entropy not as a descriptive statistic but as a control variable for allocating expensive supervision and expensive reasoning. This turns uncertainty estimation into a resource allocation problem with measurable efficiency benefits.

Aims of the work

The first aim is to formalize entropy-driven reasoning as a discipline that specifies the probabilistic object, the decision purpose, and the validation protocol. We treat this as a methodological contribution. It is also the thread that connects the five chapters.

The second aim is to determine when maximum entropy inference improves categorical distribution estimation under finite samples. This aim includes identifying regimes where MaxEnt is meaningful and regimes where it is worse than a symmetry baseline. It also includes isolating the role of constraint noise, constraint choice, and the numerical encoding used to define moments.

The third aim is to reinterpret nonnegative matrix factorization through a probabilistic common-cause model. The goal is not only reconstruction but also interpretability and stability. The work aims to derive an effective rank selection criterion grounded in predictability and to quantify stability under weak noise and random initialization.

The fourth aim is to validate next-token entropy as an uncertainty signal for multiple-choice question answering. The work aims to separate knowledge-dominated questions from reasoning-dominated questions, then to test entropy as an error discriminator and to examine calibration behavior across regimes.

The fifth aim is to convert validated entropy measurements into adaptive compute policies. This includes a fine-tuning policy that selectively applies costly chain-of-thought (CoT) distillation to high-entropy examples. It also includes an inference policy that separates reasoning generation from answer generation, then studies how model size and reasoning trace quality interact.

Scientific novelty

We contribute novelty through validated claims and operational tools. It proposes a risk-based criterion for when the maximum-entropy (MaxEnt) estimator is beneficial. The estimator is evaluated under forward KL risk, with averaging over sampling noise and over a family of priors. A concrete symmetry baseline is used, which turns the question of usefulness into a measurable comparison.

We treate representation dependence as a first-class boundary for MaxEnt. Moment-constraint MaxEnt depends on the numerical encoding of categories. This dependence is not a minor detail. It can dominate finite-sample performance. We demonstrate that misaligned encodings can render MaxEnt worse than the uniform baseline even when other estimators remain robust.

In representation learning, we contribute a predictability-grounded effective rank for NMF that is motivated by a common-cause interpretation. It also supplies a stability framework that measures how bases vary across seeds and noise. The result is a rank selection principle that is tied to interpretability and reproducibility rather than to reconstruction error alone.

We formulate directional entropy diagnostics for approximate factorization. It then tests these diagnostics empirically. This gives entropy a role as a falsifiable description of how factorization redistributes structure between images and basis elements.

In LLM evaluation, we provide a conditional validity statement for token entropy as uncertainty.

t shows that entropy can be a strong error discriminator in knowledge-dominated regimes, while ecoming weaker in reasoning-dominated regimes. This result supports regime-aware uncertainty valuation instead of universal confidence claims.

Finally, we use entropy to allocate compute in training and inference. In fine-tuning, entropy ates which examples receive chain-of-thought distillation. In inference, reasoning traces are gen-rated by a thinker model and consumed by an answerer model. We measure how performance epends on the thinker and how weak reasoning traces can harm even strong answerers.

Practical and theoretical significance

he practical significance is that we provide clear guidance on when entropy-based methods help nd when they fail. Inference practitioners can use the MaxEnt results to avoid overconfident struc-ure injection in sparse regimes and to understand when encoding choices create hidden assump-ons. Representation learning practitioners can use the NMF rank and stability tools to produce arts-based decompositions that are reproducible. LLM practitioners can use the uncertainty find-ngs to avoid naive entropy thresholding in reasoning-heavy tasks. They can also use the compute llocation ideas to build pipelines that achieve better accuracy per token.

The theoretical significance is that the we connect entropy and KL-based ideas across domains hrough a shared validation framework. We also highlighte the role of representation in defining hat entropy measures. Inference, factorization, and language modeling all depend on how states re encoded. We show that this dependence is not peripheral. It is central to whether entropy-based easoning is meaningful.

Content of the work

ntroduction

Ve begin with a survey of entropy across information theory, inference, classical ML, and LLM ractice. We emphasize that entropy is a functional of a distribution and of a sample space. It is ot a property of a single observation. This distinction matters because many practical systems ompute entropy from one model output and then treat the result as a universal confidence score. Ve argue that such a use is defensible only when validated for the target decision.

We also emphasize comparability. For LLMs, next-token entropy can be computed from the raw ogits distribution. It can also be computed after temperature scaling or truncation. These choices hange the object. As a result, two entropy values can be incomparable if they are computed under ifferent decoding rules. We therefore fix the entropy object when making comparisons and varies only when explicitly stated.

We then motivate the object–purpose–validation discipline with examples. In MaxEnt infer-nce, the object is an estimated categorical distribution. The purpose is minimizing risk under a osen loss. Validation is performed through expected risk and a symmetry baseline. In NMF, the bject is a probabilistic image distribution and a latent-factor decomposition. The purpose is inter-retable stable representations. Validation includes stability across seeds, noise and predictability onstraints. In LLM chapters, the object is the next-token distribution or the distribution along a rea-oning trace. The purpose is either error prediction or compute allocation. Validation is performed y stratified error discrimination and by measured accuracy–token trade-offs.

Chapter 1: Validity limits of the maximum entropy method

This chapter asks when the maximum-entropy (MaxEnt) principle provides a useful estimate of an unknown discrete distribution from a finite sample. Let Z take n ordered outcomes $z_1 < \cdots < z_n$ with unknown probabilities $q = \{q_k\}_{k=1}^n$, $q_k \equiv q(Z = z_k)$. From a sample of length M we observe counts $\{m_k\}_{k=1}^n$ with $\sum_k m_k = M$ and we compare estimators $\hat{q} = \{\hat{q}_k\}$.

We measure performance using the forward Kullback–Leibler loss

$$K[q, \hat{q}] = \sum_{k=1}^{n} q_k \ln \frac{q_k}{\hat{q}_k}. \tag{1}$$

We average this loss over samples drawn from q and then over a prior on q. This gives the Bayes risk $\langle \overline{K} \rangle$.

A baseline is the data-free MaxEnt solution, which is uniform,

$$q_k^{[0]} = \frac{1}{n}. \tag{2}$$

An estimator is called meaningless in a regime if it performs worse on average than $q^{[0]}$, that is $\langle \overline{K[q, \hat{q}]} \rangle > \langle \overline{K[q, q^{[0]}]} \rangle$. In that case, using data through the estimator is worse than ignoring the data.

We focus on three estimator families. Regularized maximum likelihood uses

$$p_{\mathrm{ML}}(z_k) = \frac{m_k + b}{M + nb}, \tag{3}$$

where $b \geq 0$ shrinks toward uniformity. Bayesian posterior means under Dirichlet priors have the form $p(z_k) = (m_k + \alpha_k)/(M + A)$ with $A = \sum_k \alpha_k$, and mixtures of Dirichlets yield Bayes-optimal estimators for the KL loss while becoming more sensitive to prior mismatch. MaxEnt with a first-moment constraint fixes the empirical mean $\mu_1 = \frac{1}{M} \sum_{u=1}^{M} z_{i_u}$ and gives the Gibbs form

$$q^{[1]}(z_k) = \frac{e^{-\beta z_k}}{\sum_{l=1}^{n} e^{-\beta z_l}}, \qquad \beta \text{ chosen so that } \sum_k q^{[1]}(z_k)\, z_k = \mu_1. \tag{4}$$

We also use a shrunk variant

$$q_\xi^{[1]}(z_k) = \frac{e^{-\xi\beta(z_k - \mu_1)}}{\sum_{l=1}^{n} e^{-\xi\beta(z_l - \mu_1)}}, \qquad 0 < \xi < 1, \tag{5}$$

where ξ is selected using prior information to reduce $\langle \overline{K} \rangle$.

The main point is simple. In sparse samples, MaxEnt with empirically fitted constraints can overfit unless the prior supports the structure implied by those constraints. When the prior does carry compatible structure, MaxEnt can become highly competitive. A key example is a prior that preserves $\langle q_k \rangle = 1/n$ but introduces conditional rank correlations between the ordering of z_k and the ordering of q_k.

M	$\langle \overline{K}_{\text{Bayes}} \rangle$	$\langle \overline{\mathbf{K}}_{\text{Bayes}} \rangle$	$\langle \overline{K}_{\text{ML}} \rangle_{b=b_{\text{opt}}}$	$\langle \overline{K}_{\text{ML}} \rangle_{b=1}$	$\langle \overline{K}_1 \rangle$	$\langle \overline{K}_{1\,\xi} \rangle$ (ξ_{opt})
35	0.014	0.206	0.180	0.204	0.048	0.047 (0.91)
25	0.015	0.207	0.188	0.210	0.053	0.051 (0.87)
15	0.017	0.209	0.197	0.214	0.065	0.060 (0.84)
11	0.022	0.209	0.201	0.215	0.077	0.069 (0.78)
7	0.035	0.209	0.205	0.215	0.105	0.084 (0.69)
5	0.052	0.210	0.207	0.214	0.141	0.101 (0.59)
3	0.083	0.210	0.209	0.213	0.268	0.140 (0.50)
1	0.150	0.211	0.211	0.212	—	— (—)

Table 1: $n = 60$ with $z_k = k$. The prior is a two-component Dirichlet mixture that preserves $\langle q_k \rangle = 1/n$ and encodes conditional rank correlations, with $\alpha_0 = 0.3$ and $\epsilon = 1.1$. The data-free baseline $\langle \overline{K[q, q^{[0]}]} \rangle$ equals 0.212. Values above 0.212 are worse than ignoring data. Columns show the Bayes-optimal estimator for the mixture prior, a misspecified Bayes estimator that collapses the mixture to one Dirichlet, regularized ML with $b = b_{\text{opt}}$ and with $b = 1$, MaxEnt with the first-moment constraint, and shrunk MaxEnt with ξ_{opt} in brackets. Averages are computed numerically by sampling 10^2 probability vectors and, for each, 10^2 samples of length M. For $M = 1$, the first-moment MaxEnt estimator can be degenerate for extreme samples, so those entries are omitted.

Table 1 shows this regime. The prior is a two-component Dirichlet mixture. One component ncourages $q_1 < \cdots < q_n$ and the other encourages $q_1 > \cdots > q_n$, each with probability $1/2$. In his setting, the first-moment constraint is informative because the empirical mean carries ordering nformation. MaxEnt can then outperform frequency-based estimators that do not use that ordering.

The table supports three conclusions. First, when the prior contains the ordering structure, first-noment MaxEnt can beat even optimally regularized ML for $M \geq 5$ and can remain well below he baseline 0.212 in the sparse regime $M < n$. Second, Bayesian procedures can fail badly under rior misspecification. The collapsed single-Dirichlet Bayes estimator sits close to the baseline and erforms far worse than MaxEnt for $M \geq 5$ in this experiment. Third, shrinkage of the MaxEnt onstraint can help at small M, but it requires additional prior input through ξ.

Overall, MaxEnt is not a universally safe default in finite samples. In sparse data, it is meanngful when the constraints are supported by genuine prior structure. When that support is absent, IaxEnt can overfit and lose to the uniform baseline. When it is present, as in Table 1, MaxEnt can e competitive with and sometimes superior to standard frequency-based estimators.

Chapter 2: Nonnegative matrix factorization and the principle of the common cause

his chapter studies nonnegative matrix factorization (NMF),

$$P_{\pi i} \approx \hat{P}_{\pi i} = \sum_{b=1}^{R} B_{\pi b} W_{bi}, \qquad B, W \geq 0,$$

and develops a probabilistic interpretation that links NMF to a common-cause model. The starting point is to treat each image column i as a probability distribution over pixels. After normalization, the pixel index π becomes a discrete variable and the image index i becomes another discrete variable, so the dataset defines a joint distribution $p(\pi, i)$. NMF can then be read as an approximate latent-variable factorization,

$$p(\pi, i) \approx \sum_{b=1}^{R} p(\pi \mid b)\, p(i \mid b)\, p(b),$$

where b indexes latent basis components. In this view, b plays the role of a (candidate) common-cause variable that explains dependence between π and i through conditional independence structure in the sense of the Principle of the Common Cause (PCC). The chapter does not claim that NMF recovers causal truth; rather, PCC is used as a modeling lens that yields testable, operational criteria for selecting and validating decompositions.

A key contribution is an operational rank-selection criterion grounded in predictability. In an exact common-cause factorization, conditioning on the latent state often yields sharper predictions than conditioning on the original variable. In the approximate setting, the chapter defines an effective rank R_c as the smallest rank at which this predictability property holds broadly across pixel–image relationships, up to a controlled tolerance that allows a small fraction of violations in noisy data. This yields a practical $R(\xi, \tau)$ notion that remains meaningful under perturbations, unlike criteria that can fail in weak-noise regimes.

Stability is treated as a central validity requirement for interpretability. Because NMF is nonconvex and nonidentifiable, different runs can yield different bases even when reconstruction error is similar. The chapter therefore evaluates basis reproducibility across random seeds and across noise realizations, and it makes this assessment explicit with a matching-based stability test. Two NMF runs produce two sets of basis images $\{B_a\}_{a=1}^{R}$ and $\{B_b^{[\text{noisy}]}\}_{b=1}^{R}$. Similarity is measured by cosine distance between vectorized basis images, and components are paired by solving a linear assignment problem that minimizes the total matched distance across all pairs. Figure 1 illustrates the resulting matched pairs near the effective rank: the left basis image in each pair is learned from a clean half of UTKFace, while the right basis image is learned from an independently trained, noise-corrupted half. Even when some pairs are not extremely close by cosine distance, they frequently match semantically by highlighting the same facial part. This qualitative agreement supports the operational claim: near R_c, the representation is reproducible in the sense that it recovers similar parts under weak-to-moderate perturbations, whereas at substantially larger ranks stability degrades as the factorization begins to overfit idiosyncrasies and noise.

The chapter also uses entropy as a directional diagnostic of how the approximation reshapes distributional structure. For each image i, the normalized distribution $p(\pi \mid i)$ has entropy

$$S_i = -\sum_{\pi} p(\pi \mid i) \ln p(\pi \mid i).$$

The NMF approximation induces $\hat{p}(\pi \mid i)$ and thus $\hat{S}_i$. Empirically, approximate factorization often smooths the per-image pixel distributions, yielding $S_i \leq \hat{S}_i$ in typical cases. This is not presented as a universal law; it is treated as an observable signature of how the approximation redistributes mass (often blurring sharp pixel-level structure while retaining higher-level parts).

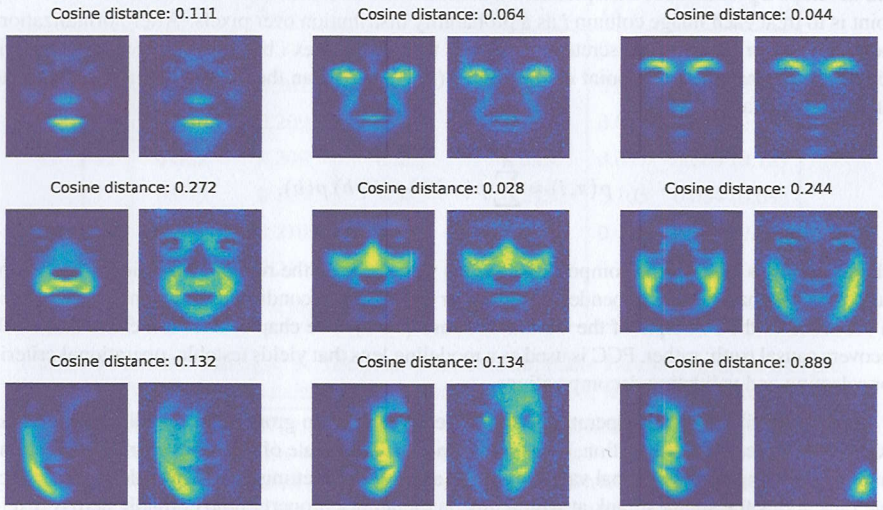

Figure 1: Illustration of basis-image stability via matching under perturbations (UTKFace dataset). Each row shows an optimally matched pair of NMF basis images at rank $R = 36$. Left: a basis image learned from the first half of the dataset. Right: the matched basis image learned from the second half after applying symmetric pixel-flip noise with probability $\xi = 0.25$ (and using a different optimization seed). Pairs are obtained by solving a global assignment that minimizes the total cosine distance between basis vectors across the two runs. Many matched pairs remain semantically aligned (e.g., emphasizing similar facial regions), illustrating reproducible parts-based structure near the effective rank.

At the same time, basis-level distributions $p(\pi \mid b)$ often become more localized for ranks in he stable regime, which corresponds to lower basis entropies and provides a quantitative reflection f the parts-based property that motivates NMF interpretability.

Chapter 3: When an LLM is apprehensive about its answers and when its uncertainty is justified

his chapter examines token entropy as an uncertainty signal for multiple-choice question answering (QA). The operational questions are: when the model's answer distribution is diffuse, does rror probability increase, and when it is sharp, does correctness increase? The chapter evaluates oth discrimination (ranking errors by entropy) and calibration (mapping entropy to correctness robabilities).

The object is the model's next-token distribution at answer time. Entropy is computed from the ll vocabulary distribution. This choice aims to capture model belief rather than the behavior of a ampling rule. The chapter emphasizes that entropy can be distorted by decoding transformations. therefore fixes the definition and uses it consistently in evaluation.

A key methodological feature is stratification. Multiple-choice QA mixes failures caused by missing knowledge and failures caused by faulty reasoning. The chapter introduces an automated annotation pipeline, using model-as-judge, to label questions by whether they are primarily knowledge-dominated or reasoning-dominated. It also approximates reasoning burden through step-like heuristics and structured judgments. The chapter treats these labels as imperfect. It uses them for stratification rather than as ground truth.

The main result is conditional validity. In knowledge-dominated regimes, entropy tends to be higher for incorrect answers than for correct answers. This yields useful discrimination. In reasoning-dominated regimes, entropy becomes weaker as a predictor. The model can be confident and wrong. It can assign high probability to a wrong option because the internal reasoning is flawed or because it anchors early on a misleading pattern. In that case, the output distribution remains sharp even though the answer is incorrect. The chapter presents this as a boundary. Entropy reflects uncertainty about competing surface completions. It does not necessarily reflect uncertainty about whether the internal reasoning is valid.

Calibration is assessed separately: even when entropy separates errors well, mapping entropy to a calibrated probability of correctness can be biased. We frame this as an engineering concern. Systems that use entropy for abstention or escalation need calibration checks, not only ranking checks.

The chapter also reports negative findings. Model-as-judge scores, in the tested form, do not serve as robust correctness predictors. This reinforces this work's methodological stance. Plausible uncertainty proxies must be validated for the target regime and the target decision. They cannot be adopted because they sound aligned with intuition.

This chapter motivates later chapters. If answer-time entropy is not reliable for reasoning-heavy questions, then it becomes necessary to either improve reasoning traces or to build new signals that evaluate reasoning quality. That is precisely what the compute-aware chapters explore, by treating reasoning as an allocatable resource and by separating thinking from answering.

Chapter 4: Complexity-aware fine-tuning

This chapter studies domain adaptation under explicit compute budgets. It focuses on compact LLMs, where supervised fine-tuning is feasible, but large-scale distillation is still expensive because it requires long teacher generations and substantially increases the number of training tokens. The core question is therefore not only how to improve accuracy, but how to do so while controlling the token budget.

The chapter proposes a selective supervision policy that uses entropy as an automatic proxy for example difficulty. The entropy object is defined at answer time under an answer-only prompt: the student model is asked to output only the option label (one token), and we compute the full-vocabulary next-token entropy of that answer step,

$$S = -\sum_{v \in \mathcal{V}} p_\theta(v \mid \text{prompt}) \log p_\theta(v \mid \text{prompt}).$$

High entropy means that probability mass is spread across multiple plausible options, which empirically correlates with higher error risk and indicates that richer supervision may have higher marginal value.

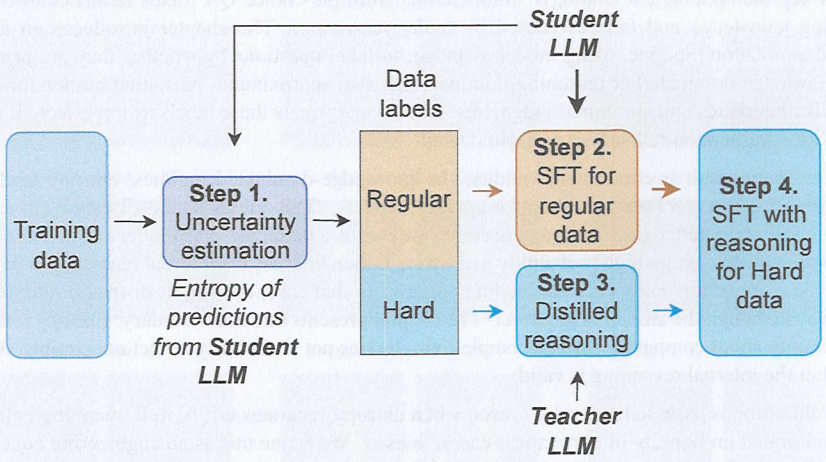

Figure 2: Complexity-aware fine-tuning scheme for a student LLM. Step 1: estimate question complexity via the student's uncertainty, measured by single-token answer entropy under an answer-only prompt. Step 2: apply vanilla SFT on regular-complexity data. Step 3: for hard data, elicit chain-of-thought from a stronger teacher and attach it to the training set. Step 4: fine-tune the student on the reasoning-enriched hard subset. The central idea is selective allocation of expensive reasoning supervision: distillation is used only where the student is most uncertain.

Figure 2 summarizes the resulting pipeline. First, we score each training example by the student's single-token answer entropy (Step 1). We then split the dataset into a regular band (low-to-moderate entropy) and a hard band (high entropy). For the regular band we apply standard supervised fine-tuning (SFT), because the model already has a strong preference for one option and additional reasoning traces are often redundant (Step 2). For the hard band we invoke a stronger teacher model to generate a CoT trace, and we distill that reasoning into the student by fine-tuning on reasoning-enriched examples (Steps 3–4). Operationally, the policy replaces a fragile "train-longer stage-by-epoch" curriculum with a fixed, entropy-defined routing rule: the expensive reasoning supervision is tied to difficulty signals at the example level rather than to a hand-tuned training schedule.

The chapter validates this policy against multiple baselines: plain SFT on all data, full chain-of-thought distillation on all data, curriculum-style SFT, and alternative allocations that apply distillation to the wrong subsets. The reported experiments show a strong improvement in the accuracy–token frontier. After 10 epochs in the studied setting, a Qwen 3B student reaches about 0.42 accuracy with about $29k$ processed tokens under the SFT baseline, while full distillation reaches about 0.49 with about $19.72k$ tokens. The entropy-gated pipeline reaches about 0.50 with about $3.99k$ tokens. For Phi-4-mini, SFT reaches about 0.55 with about $27k$ tokens, full distillation reaches about 0.63 with about $15.15k$ tokens, and the pipeline reaches about 0.60 with about $2.67k$ tokens. With longer training (20 epochs), the pipeline continues improving while retaining large token savings relative to full distillation (e.g., Qwen 3B around 0.52 at $7.98k$ tokens and Phi-4-mini around 0.64 at $5.35k$ tokens in the reported runs).

These results support the chapter's interpretation in terms of marginal value of compute. Distillation provides the greatest benefit where the student is uncertain and likely to be wrong; where the student is already confident, chain-of-thought supervision can spend many tokens while adding little signal. In this chapter, entropy is therefore not used as a generic description of "uncertainty," but as a validated control variable for allocating expensive supervision under a measured accuracy–compute trade-off.

Chapter 5: Better thinking or a bigger model: thinking and answering shuffles with Qwen3 on GPQA

This chapter studies a modular inference design. Reasoning is generated by a thinker model. The final answer is produced by an answerer model that conditions on the reasoning trace. The central question is whether performance is determined mainly by the size of the answerer or by the quality of the reasoning trace. A second question follows. If the thinker dominates, then strong thinking traces could be reused and amortized across cheaper answerers.

This chapter evaluates this idea using a single model family to ensure compatibility. Five Qwen3 dense sizes are used (where B denotes billions of parameters): 0.6B, 1.7B, 4B, 8B, and 14B. The evaluation is conducted on the GPQA-main split (448 questions). For each pairing, the thinker generates a chain-of-thought trace. The answerer then outputs only the final choice conditioned on that trace. All 25 pairings are evaluated. Decoding is deterministic to isolate capability and trace effects rather than sampling variation.

The chapter records chain-level statistics. Thinking length is measured as the number of tokens in the reasoning trace. Thinking entropy is computed as the mean next-token entropy over the thinking segment. At each step of the trace, entropy is computed on the full distribution. These values are then averaged along the trace. This produces a measure of how diffuse the model's next-step beliefs are while it is reasoning.

The results show a strong asymmetry that supports the dominance of the thinker. The best diagonal pairing is 14B → 14B at 59.15%. A strong thinker paired with a weak answerer remains strong. The pairing 14B → 0.6B reaches 54.24%. A weak thinker paired with a strong answerer is poor. The pairing 0.6B → 14B reaches 20.54%. This directionality is also visible in row and column means. Varying the thinker produces large accuracy shifts. Varying the answerer produces smaller shifts.

The chapter provides a mechanism-level explanation rooted in autoregressive conditioning. The answerer is not a free solver. It is conditioned on a prefix. If the prefix contains a wrong hypothesis or a misleading chain, the answerer can be anchored to that trajectory. Long traces can also create context competition, where relevant evidence is diluted by irrelevant intermediate text. In addition, once a trace commits to a wrong plan, the answerer can inherit constraints that steer decoding away from the correct option. These effects explain why weak traces can poison even strong answerers. Strong answerers are not truth filters when forced to condition on misleading scaffolds.

Thinking length patterns also align with the story. Average thinking length decreases dramatically as thinker size increases. Reported values go from around 14566 tokens at 0.6B to around 4639 tokens at 14B. This suggests that stronger thinkers reach conclusions more efficiently and with less wandering text. Mean thinking entropy is not monotonic across sizes, but the strongest thinker

nows relatively high mean entropy, with a reported example around 0.416 at 14B. The chapter nterprets this carefully. Higher entropy during thinking does not guarantee better reasoning. The veakest thinker can also show relatively high entropy while performing poorly. The important oint is that successful reasoning can maintain non-degenerate uncertainty while still converging n correct structure. Entropy must be interpreted together with outcomes and with trace quality.

The compute implication is direct. If a strong thinker dominates, then a practical system can enerate high-quality traces with a strong model, then pair them with smaller answerers for low-atency deployment. Traces could also be cached for repeated question types or for similar prompts. lowever, the chapter also shows a safety boundary. Using weak traces is harmful. Therefore, reuse nd routing policies need trace quality control. Entropy along the trace, trace length, and other nternal signals can become auditing features. We treat this as a direction for building safe modular ipelines.

This chapter completes the work's compute-allocation narrative. Chapter 4 allocates training-me compute by gating expensive supervision. Chapter 5 allocates inference-time compute by eparating thinking from answering and by showing where model size matters most.

Conclusion and outlook

ntropy is a fundamental tool for quantifying uncertainty and structure in AI systems, but its use-ulness depends on clear definitions and careful validation. When applied with a well-specified bject, explicit purpose, and comparison to meaningful baselines, entropy provides actionable in-ights across inference, representation learning, model evaluation, and compute-aware adaptation. Our work demonstrates both the strengths and the limitations of entropy in these contexts.

1. **Operational discipline.** Entropy becomes informative only under an *object–purpose–validation* discipline: specify the sample space and distribution, specify the decision objective, and validate against outcomes and a baseline representing minimal unsupported structure.

2. **Finite-sample inference limits.** Maximum-entropy (MaxEnt) inference has sharp validity limits in sparse data: noisy constraints and near-uniform true worlds can make MaxEnt inject spurious structure and increase forward KL risk, rendering it operationally meaningless under a symmetry baseline; conversely, aligned prior structure and matched constraint forms can make MaxEnt beneficial.

3. **Representation dependence is central.** Representation choices are not peripheral: in MaxEnt, numerical encodings defining moment constraints encode assumptions that can dominate performance and determine what entropy-based reasoning actually measures.

4. **Interpretable factorization via common cause.** Nonnegative matrix factorization (NMF) can be interpreted as an approximate common-cause factorization of a joint distribution over pixels and images; this motivates an effective-rank criterion based on predictability and elevates stability testing as a necessary condition for interpretability, with entropy serving as a directional diagnostic of smoothing at the image level and localization at the basis level.

5. **Conditional validity of LLM entropy.** Next-token entropy is validated as an uncertainty signal in knowledge-dominated multiple-choice QA, but becomes weaker in reasoning-dominated regimes where confident wrong answers are common; hence, naive global entropy thresholding can succeed in some tasks and fail in others.

6. **Compute-aware reliability.** Validated entropy signals can drive resource allocation: entropy-gated chain-of-thought distillation improves the accuracy–token frontier in fine-tuning, and modular thinker–answerer inference shows that reasoning quality dominates answerer size (strong thinkers enable small answerers; weak traces can poison strong answerers), motivating trace quality control.

Future directions include decision-theoretic policy evaluation under explicit cost models (abstention, escalation, selective computation), reasoning-aware uncertainty signals computed along trajectories (entropy paths, step consistency, learned trace-validity detectors), learned routing optimized for accuracy–latency under distribution shift, broader replication of thinker–answerer effects across model families and tasks, and treating representation choices as design variables across inference, factorization, and language modeling (encodings, normalization, token spaces, and decoding rules).

Publications in the topic of thesis

(1) A. E. Allahverdyan, E. A. Khalafyan, and N. H. Martirosyan, "Validity limits of the maximum entropy method," *Chinese Journal of Physics*, vol. 71, pp. 95–111, 2021.

(2) E. Khalafyan, A. Allahverdyan, and A. Hovhannisyan, "Nonnegative matrix factorization and the principle of the common cause," *2025 IEEE 12th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 1–10, 2025.

(3) P. Sychev, A. Goncharov, D. Vyazhev, E. Khalafyan, and A. Zaytsev, "When an LLM is apprehensive about its answers–and when its uncertainty is justified," *Zapiski Nauchnykh Seminarov POMI*, 2026.

(4) A. Goncharov, D. Vyazhev, P. Sychev, E. Khalafyan, and A. Zaytsev, "Complexity-aware fine-tuning," *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics*, 2026.

(5) E. A. Khalafyan, "Better Thinking or a Bigger Model? Thinking–Answering Shuffles with Qwen3 on GPQA," *Mathematical Problems of Computer Science*, vol. 64, pp. 17–28, 2025.

Խալաֆյան Էդվարդ Արսենի

Էնտրոպիայով առաջնորդվող ԱԲ. հավանականային եզրահանգում, պատճառահետևանքային ներկայացումներ և մոդելների ադապտիվ ճշգրտում

Ամփոփագիր

Էնտրոպիան առանցքային հասկացություն է վիճակագրության, մեքենայական ուսուցման և ժամանակակից մեծ լեզվական մոդելների մշակման ոլորտներում: Այն նութագրում է պատահական մեծության կամ ինֆորմացիայի աղբյուրի անորոշության ափը, տվյալների հաղորդման ու սեղմման սահմանները, պատկերացում ոալիս համակարգում կարգավորվածության աստիճանի մասին, ինչպես նաև ահմանափակ գիտելիքի պայմաններում եզրահանգումների վստահելիության վերաբերյալ: Կիրառական խնդիրներում այն հաճախ օգտագործվում է բաշխումների վերականգնման համար և որպես կառավարման ազդանշան՝ կանխատեսման վստահելիությունը կարգավորելու միջոց: Միևնույն ժամանակ, էնտրոպիան հաճախ կիրառում են որպես ունիվերսալ մեծություն, առանց ստուգելու նման կիրառման ռեկտությունը, անհրաժեշտությունն ու կայունությունը՝ տվյալների կամ ուսուցման պայմանների փոփոխության դեպքում:

Այս աշխատանքում առաջարկվում է վավերականության վրա կենտրոնացած ոտեցում՝ էնտրոպիայով կառավարվող ԱԲ-ի համար: Մենք էնտրոպիան դիտարկում նք որպես գործառնական մեծություն միայն այն բանից հետո, երբ ստուգում ենք, ր դրա կիրառումն իրականում օգտակար է ընտրված օբյեկտի և առաջադրված նդրի շրջանակում: Օբյեկտը կարող է լինել հավանականությունների բաշխումը կամ եզվական մոդելում հաջորդ թոքենի գեներացման «վստահությունը»: Խնդիրը կարող ներառել վերականգնում, մոդելի ընտրություն, հուսալիության գնահատում կամ աշվարկային ծախսերի կառավարում: Այնուհետև մենք ընտրված կիրառման տարբերակը վավերացնում ենք դիտարկելի չափորոշիչներով, ինչպիսիք են շտությունը, կայունությունը և հաշվարկային արդյունավետությունը:

Վիճակագրական եզրահանգման խնդիրներում մենք ուսումնասիրում նք առավելագույն էնտրոպիայի սկզբունքը՝ տվյալներից գնահատվող ահմանափակումների ներքո: Մենք ստուգում ենք, երբ է այն ինֆորմատիվ վերջավոր տվյալերի դեպքում՝ որակը միջինացնելով բազմաթիվ գեներացնող աշխումների և բազմաթիվ տվյալների բազմության վրա: Սա թույլ է տալիս առանձնացնել ռեժիմներ, որտեղ առավելագույն էնտրոպիան բարելավում է որոշումների սպասվող որակը, և ռեժիմներ, որտեղ այն ներմուծում է չարդարացված կառուցվածք: Կիրառելիության սահմանները որոշվում են տվյալների ծավալով և ահմանափակումների ընտրությամբ:

Դասական մեքենայական ուսուցման մեջ մենք կենտրոնանում ենք ոչբացասական ատրիցային ֆակտորիզացիայի վրա՝ որպես ներկայացումների մեթոդ՝ ավանականային և պատճառահետևանքային մեկնաբանությամբ: Ընդհանուր պատճառի սկզբունքով մեկնաբանությունը տալիս է ռանկի ընտրության գործնական եթոդ և նպաստում է թույլ աղմուկի և տարբեր իրականացումների նկատմամբ յունության վերլուծությանը: Էնտրոպիան կիրառվում է որպես ուղղորդված սխտորոշիչ ցուցիչ՝ թե ինչպես է ֆակտորիզացիան փոխում բաշխումային

կառուցվածքը և ներկայացումների նոսրությունը: Այս ազդեցությունները ստուգվում են փորձնականորեն, այլ ոչ թե ներկայացվում որպես էվրիստիկա:

Մեծ լեզվական մոդելներում էնտրոպիան հաշվարկվում է հաջորդ թոքենի հավանականային բաշխումից: Մենք ստուգում ենք, թե երբ է թոքենի գեներացիայի էնտրոպիան կանխատեսում սխալները և երբ է այն դադարում լինել հուսալի ազդանշան: Այն օգտակար է գիտելիքահեն ռեժիմներում և ավելի քիչ հուսալի՝ բազմաքայլ դատողությամբ պայմանավորված ռեժիմներում, որտեղ հնարավոր են վստահ, բայց սխալ պատասխաններ: Սա պահանջում է ռեժիմից կախված անորոշության գնահատում և գեներացիայի հատվածային վերլուծություն:

Այնուհետև մենք օգտագործում ենք վավերացված էնտրոպիական ազդանշանները՝ հաշվարկների բաշխման համար: Կարգավորման ժամանակ էնտրոպիայի վրա հիմնված բարդության գնահատումը տվյալները բաժանում է այնպես, որ «թանկ» դատողությունները կիրառվեն միայն այնտեղ, որտեղ տալիս են շահում:

Մեծ լեզվական մոդելների դատողություն անելու ընթացքի պարագայում, դիտարկում ենք «մտածողի» և «պատասխանողի» դեկոմպոզիցիա, որտեղ մտածողը գեներացնում է դատողության հետագիծ, իսկ պատասխանողը վերադարձնում է միայն պատասխանի պիտակը: Արդյունքները ցույց են տալիս, որ այս սցենարով, մտածող խոշոր մոդելները կարող են ուժեղացնել պատասխանող փոքր մոդելների որակը, մինչդեռ թույլ մտածող մոդելները կարող են վատթարացնել նույնիսկ խոշոր պատասխանող մոդելների արդյունքը: Էնտրոպիայի և դատողության երկարության վիճակագրությունները օգնում են տարբերակել օգտակար դատողության հետագծերը մոլորեցնողներից:

Քննարկված բոլոր դոմեններում էնտրոպիան դառնում է ինֆորմատիվ միայն այն դեպքում, երբ կապվում է հստակ սահմանված օբյեկտի և դրված խնդրի հետ: Մենք դիտարկում ենք էնտրոպիան ոչ թե որպես համընդհանուր մեծություն, այլ որպես չափելի ազդանշան՝ կախված խնդրի լուծման որակից և առկա ռեսուրսներց:

Халафян Эдвард Арсенович

Энтропийно-управляемый ИИ: вероятностный инференс, причинные представления и адаптивное дообучение моделей

Резюме

Энтропия является центральной величиной в статистике, машинном обучении и современных больших языковых моделях. Она связывает кодирование и сжатие информации, физические представления о неупорядоченности и формализацию неопределенности при ограниченном знании. В прикладных задачах ее часто используют для восстановления распределений и как сигнал управления предсказательной уверенностью. При этом энтропию нередко применяют как универсальную велечину, не проверяя корректность, необходимость и устойчивость такого применения при смене данных или условий обучения.

В этой работе предлагается подход к энтропийно-управляемому ИИ, ориентированный на валидность. Мы рассматриваем энтропию как операциональную величину только после того, как проверим, что ее применение действительно полезно для выбранного объекта и в рамках поставленной задачи. Объектом может быть распределение вероятностей или "уверенность" генерации следующего токена в языковой модели. Задача может включать восстановление, выбор модели, оценку надежности или управление вычислительными затратами. Далее мы валидируем выбранный вариант применения по наблюдаемым метрикам, таким как точность, устойчивость и вычислительная эффективность.

В задачах статистического вывода мы изучаем принцип максимальной энтропии при ограничениях, оцениваемых по данным. Мы проверяем, когда он информативен в конечных выборках, усредняя качество по множеству порождающих распределений и множеству выборок. Это позволяет выделить режимы, где максимальная энтропия улучшает ожидаемое качество решений, и режимы, где она вносит необоснованную структурность. Границы применимости определяются объемом данных, выбором ограничений и выбранным представлением.

В классическом машинном обучении мы фокусируемся на неотрицательной матричной факторизации как на методе представлений с вероятностной и причинной интерпретацией. Интерпретация через принцип общей причины дает практическое правило выбора ранга и поддерживает анализ устойчивости к слабому шуму и к различным инициализациям. Энтропия используется как направленный диагностический признак того, как факторизация изменяет распределительную структуру и разреженность представлений. Эти эффекты проверяются экспериментально, а не вводятся как эвристика.

В больших языковых моделях часто вычисляется энтропия распределения вероятностей следующего токена генерации. Мы проверяем, когда энтропия на уровне токенов предсказывает ошибки и когда она перестает быть надежным индикатором. Она полезна в режимах доминирования знаний и менее надежна в режимах доминирования многошагового рассуждения, где возможны уверенные ошибки. Это требует режимно-зависимой оценки неопределенности и анализа по сегментам генерации.

Далее мы используем валидированные энтропийные сигналы для распределения вычислений. При обучении оценка сложности на основе энтропии маршрутизирует данные так, что дорогое обучение на рассуждениях применяется только там, где она дает выигрыш.

При инференсе больших языковых моделей мы рассматриваем декомпозицию на мыслителя и ответчика, где мыслитель генерирует рассуждение, а ответчик выдает метку ответа. Результаты показывают, что сильные рассуждения хорошо переносятся между ответчиками, тогда как слабые рассуждения могут ухудшать качество даже крупных ответчиков. Статистики энтропии и длины помогают отличать полезные рассуждения от вводящих в заблуждение.

Во всех рассмотренных доменах энтропия становится информативной только будучи привязанной к явно заданному объекту и задачи. Мы рассматриваем энтропию не как универсальный прокси-показатель, а как измеримый сигнал, связанный с качеством на задаче и учетом ресурсов.