

Bottom-Up Interpretability of Pretraining Dynamics via Loss Curve Decomposition

Koshiro Aoki^{1,2*} Masaru Isonuma^{2,3,4,5} Max Müller-Eberstein^{4,6} Yusuke Oda^{2,7}
 Hirokazu Kiyomaru² Takashi Kodama² Chaoran Liu² Yohei Oseki^{2,4}
 Yusuke Miyao^{2,4} Daisuke Kawahara^{1,2}
¹Waseda University ²NII LLMC ³Tohoku University ⁴The University of Tokyo
⁵RIKEN ⁶IT University of Copenhagen ⁷Nara Institute of Science and Technology
 aokikoshiro@akane.waseda.jp

Abstract

Average loss curves during language model pretraining obscure diverse training dynamics across individual training data instances. Gaining a more granular understanding of these dynamics currently requires tracking intermediate checkpoint performance on predefined tasks, or computing higher-order gradients with respect to each data instance. To instead enable a scalable bottom-up decomposition of language model training dynamics, we propose identifying data-driven *learning domains* using non-negative matrix factorization (NMF) to group training data based solely on their per-instance losses, without any gradient computation. Applied to 3.2k checkpoints across 21 configurations of the Pythia model suite, our method automatically recovers distinctive training dynamics for different training data clusters. The resulting learning domains are consistent across model sizes and random seeds, while being interpretable as, e.g., code, mathematics, repetitive patterns, and non-English texts. The corresponding domain-level loss curves further reveal potential causes for low-performing training runs that do not surface in the average loss alone.

1 Introduction

During language model pretraining, the gradual decrease of the average cross-entropy loss curve is a useful indicator for monitoring optimization, but it compresses many different learning events into a single scalar. In particular, training examples and token-in-context predictions exhibit variable learning dynamics: some are acquired early and consistently (Swayamdipta et al., 2020; Xia et al., 2023), while others show non-monotonic loss changes (Xia et al., 2023; Chang et al., 2024). Abrupt changes in aggregate loss have also been associated with capability acquisition (Chen et al., 2024). Simultaneously, linking these instance-level changes back to higher-level model behaviors requires a way to group data with similar learning dynamics together.

Existing approaches to analyzing pretraining dynamics struggle to balance this divide between top-down and bottom-up approaches. One line of work tracks a model’s higher-level behaviors (Olsson et al., 2022; Müller-Eberstein et al., 2023; Chen et al., 2024), or geometric characteristics (Cheng et al., 2025; Li et al., 2025), resulting in interpretable learning dynamics. However, these analyses require that the quantity of interest is specified in advance. Conversely, another line of work studies training dynamics at the level of individual tokens or instances (Tirumala et al., 2022; Xia et al., 2023; Chang et al., 2024), as well as by decomposing instance-level loss changes with Hessian-based methods (Kangaslahti et al., 2026). These approaches reveal instance-level variation, but are either difficult to interpret at a global level, or require gradient computation that is hard to scale across larger models.

*This work was done while the first author was at NII LLMC.

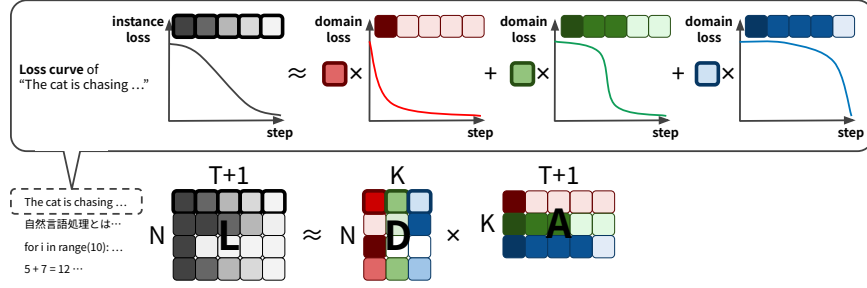


Figure 1: Overview of our loss curve decomposition. The loss matrix L , whose rows are per-instance loss curves across training checkpoints, is factorized via NMF into a membership degree matrix D and a domain loss matrix A . Each instance’s loss curve is approximated as a non-negative linear combination of K domain loss curves (the rows of A), with membership degrees given by the corresponding row of D .

To balance global interpretability and instance-level granularity while maintaining computational efficiency, we propose decomposing per-instance loss curves, as shown in Figure 1. For a fixed set of texts, we collect the loss of each instance across checkpoints and form a loss matrix whose rows correspond to per-instance loss curves. We then apply non-negative matrix factorization (NMF; Lee & Seung, 2000) to this matrix to recover a small set of domain loss curves, together with soft assignments to these domains for each instance. This decomposition uses only forward-pass losses and does not require gradients, probes, or predefined task labels. We call the resulting latent groups with shared training dynamics *learning domains*. They capture patterns of training dynamics that are shared across many instances, but are hidden within the averaged loss curve.

We apply this method to 3.2k checkpoints from 21 configurations in the Pythia model suite (Biderman et al., 2023; van der Wal et al., 2025). The recovered learning domains are interpretable as code, mathematics, repetitive patterns, non-English texts, and other categories defined by content or structure. Their loss curves differ in onset and rate of learning even when The Pile’s data source component curves appear smooth. In addition to our learning domain decomposition being stable across random seeds and model sizes, we further show how it makes an anomalous training run easier to diagnose: in Pythia-410M-seed3, the previously reported loss spike around step 75,000 consists not only of a loss *increase* for general text domains but also of a loss *decrease* for noisy text domains.

Our contributions are as follows:

- We introduce a scalable bottom-up decomposition of instance-level pretraining loss curves into learning domains without gradient computation or predefined task labels.
- We show that the recovered domains are interpretable and reveal distinct training dynamics that are neither visible in the average, nor the data source-wise, loss curves.
- We demonstrate that these domains are largely reproducible across random seeds and model sizes, and that they help localize anomalous behavior in unstable training runs.

2 Related work

Per-instance training dynamics. Several studies have shown that training dynamics vary across individual data instances. In supervised learning, training-time confidence and variability can be used to separate data instances into easy, ambiguous, and hard regions (Swayamdipta et al., 2020), while forgetting events (Toneva et al., 2019) show that some instances are repeatedly unlearned during training. In self-supervised LM pretraining, token-level learning curves show consistent acquisition order and comparable forgetting patterns across runs (Chang et al., 2024). Across model sizes, memorization has been shown to increase for larger models, while avoiding classical overfitting (Tirumala et al., 2022). In parallel, similar token subsets were found to account for the majority of loss reduction across model sizes (Xia et al., 2023). Motivated by such differences in training dynamics across

data instances, we ask which texts share similar loss trajectories and whether those shared trajectories can be summarized as a small number of interpretable latent domains.

Top-down analysis of training dynamics. Studies of in-context learning (Olsson et al., 2022), the acquisition of syntactic (Chen et al., 2024) and related linguistic capabilities (Müller-Eberstein et al., 2023), word-level shifts from unigram frequency to higher-order n -gram statistics (Michaelov et al., 2025), and representation geometry (Cheng et al., 2025; Li et al., 2025) have described transitions during LM training that are not visible in average loss curves. These approaches are top-down: i.e., each requires selecting the observable of interest in advance, and formalizing it as a probing methodology. Our method aims to uncover similar mismatches between average loss and hidden learning dynamics, but recovers them directly from the training data itself without specifying the target task beforehand.

Bottom-up analysis of training dynamics. Prior work has examined per-instance loss changes along the eigenvectors of the training loss Hessian to identify hidden phase transitions (Kangaslahti et al., 2026). However, it requires gradient computations at each checkpoint and has been demonstrated on models with up to 40M parameters. While instance-wise losses have also been used to compare models (Oyama et al., 2025) and their training trajectories (Kishino et al., 2026), these methods compare models at a global level rather than decomposing instances into domains with shared learning dynamics. Our method combines the scalability of zeroth-order statistics with a decomposition of per-instance loss curves, yielding soft domain assignments that are interpretable and reproducible across models.

3 Method

Motivated by the quantization hypothesis (Michaud et al., 2023; Michaud, 2026), we model the loss change of each instance as a superposition of loss changes associated with the skills involved in processing it. Intuitively, if the processing of two texts depends on similar skills, their losses should change in similar ways. Simultaneously, most texts do not rely on any single skill but rather follow a mixture of underlying training dynamics. Following this view, we first construct a loss matrix by stacking per-instance loss curves before decomposing this matrix such that each individual learning trajectory can be written as a sum over a small set of shared domain loss curves (Figure 1). This decomposition thereby identifies groups of texts with similar training dynamics, without requiring predefined task labels or gradient computations, yielding bottom-up *learning domains*.

Loss matrix. Consider N text instances evaluated at $T + 1$ training checkpoints (steps $t = 0, 1, \dots, T$). Let $\ell_i^{(t)} \geq 0$ denote the loss of instance i at step t . We arrange these values into a loss matrix:

$$L = [\ell^{(0)}, \dots, \ell^{(T)}] \in \mathbb{R}_{\geq 0}^{N \times (T+1)} \quad (1)$$

where each row $\ell_i = (\ell_i^{(0)}, \dots, \ell_i^{(T)})$ is the loss curve of instance i across training.

Loss curve decomposition. While instance-wise loss curves could simply be clustered via, e.g., K-means (MacQueen, 1967), a hard cluster assignment would fail to capture that many instances may depend on a mix of underlying latent loss changes. We therefore first decompose L into a small number of shared domain loss curves, where each instance’s loss curve is modeled as an additive combination of K domain loss curves. This decomposition can be written as non-negative matrix factorization (NMF; Lee & Seung, 2000):

$$L = DA + E \quad (2)$$

where $D \in \mathbb{R}_{\geq 0}^{N \times K}$ is the membership degree matrix, $A \in \mathbb{R}_{\geq 0}^{K \times (T+1)}$ is the domain loss matrix, and $E \in \mathbb{R}^{N \times (T+1)}$ is the residual matrix. Each entry $D_{i,k}$ gives the membership degree of instance i in learning domain k , and each row of A is the loss curve of domain k . Our goal is to estimate D and A so that a small number of domains explains the total variation in per-instance loss curves. NMF matches this setting because both the observed loss matrix and the desired additive components are non-negative.

Optimization problem. We estimate D and A by minimizing the Frobenius norm of the residual matrix E . According to the quantization hypothesis (Michaud et al., 2023; Michaud, 2026), skill acquisition is discrete, so changes in domain loss across checkpoints are expected to be sparse. We therefore regularize A with total variation (TV; Rudin et al., 1992) along training steps, yielding the following optimization problem:

$$\begin{aligned} \min_{D \geq 0, A \geq 0} \quad & \frac{1}{2} \|L - DA\|_F^2 + \lambda \text{TV}(A) \\ \text{s.t.} \quad & \sum_{k=1}^K D_{i,k} = 1 \quad \forall i = 1, \dots, N \end{aligned} \quad (3)$$

$$\text{TV}(A) = \sum_{k=1}^K \sum_{t=1}^T |A_{k,t} - A_{k,t-1}| \quad (4)$$

where λ is the regularization coefficient of TV. When $\lambda = 0$, the objective reduces to standard NMF with row-normalized memberships. The simplex constraint makes each row of D interpretable as a distribution over domains. We solve this optimization problem (Eq. 3) with extended multiplicative update rules (details in Appendix D).

Interpreting learning domains. To interpret each learning domain, we inspect representative texts both manually and with an LLM. For each domain k , we select the instances with the highest membership degrees $D_{i,k}$ and examine their content. This provides a descriptive label for each domain without using predefined categories.

Aligning domain indices across models. The ordering of domains from NMF is non-deterministic. To compare domain decompositions across experiments without having to manually re-label each learning domain, we therefore align all domains with respect to a single reference with pre-interpreted domain labels. In our experiments, we thus first apply labels to the learning domains from Pythia-160M (see Table 1), and then find the permutation π^* for each other model configuration that maximizes the total Pearson correlation of the domain losses and membership degrees:

$$\pi^* = \arg \max_{\pi} \sum_{k=1}^K [\rho(a_k^{\text{ref}}, a_{\pi(k)}^{\text{tgt}}) + \rho(d_k^{\text{ref}}, d_{\pi(k)}^{\text{tgt}})] \quad (5)$$

where a_k is the k -th row of A (a domain loss curve), d_k is the k -th column of D (membership degrees across all instances), and ρ denotes Pearson correlation. This procedure allows for direct comparisons of training dynamics across random seeds and model sizes.

4 Experimental settings

Models. In our experiments, we analyze six model sizes of the Pythia model suite (Biderman et al., 2023) with 70M, 160M, 410M, 1B, 1.4B, and 6.9B parameters, respectively. The models were trained on The Pile (Gao et al., 2020) with the same data ordering across model sizes. To assess the stability of our domain decompositions across random initializations, we further include experiments using the PolyPythias (van der Wal et al., 2025), covering four seeds each for the 70M, 160M, and 410M models (with variation in data ordering and weight initialization). For 160M, we also analyze six variants in which only the data ordering or the weight initialization is varied separately. This yields a total of 21 main model configurations. As a supplementary cross-family experiment, we also analyze 01mo-3-1025-7B from the OLMo 3 model family (Olmo et al., 2025).

Checkpoints. For each Pythia model, we use 154 training checkpoints spanning the full training trajectory: the random initialization (step 0), ten log-spaced checkpoints at steps 1, 2, 4, ..., 512, and 143 linearly spaced checkpoints at steps 1,000, 2,000, ..., 143,000. For OLMo 3, we use 151 checkpoints from Stage 1 pretraining: every 1,000 steps from step 0 through 10,000 and every 10,000 steps thereafter through 1,410,000.

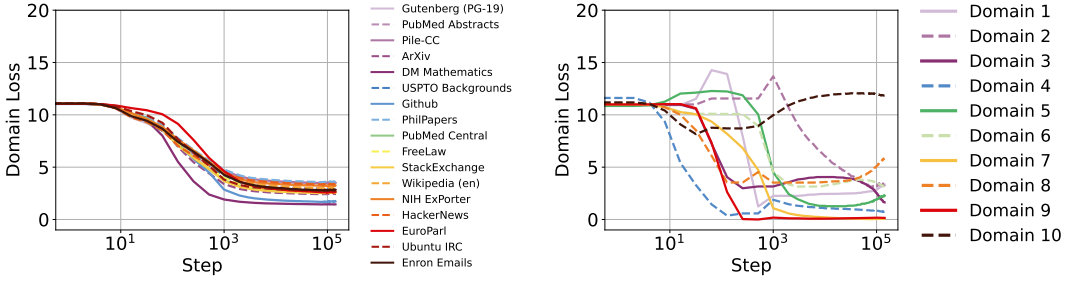


Figure 2: Loss curves for Pythia-160M. **Left:** Average loss curves of The Pile’s data sources, with each source component curve decreasing in a similar manner. **Right:** Domain loss curves from our NMF decomposition ($K=10, \lambda=1$), showing distinct training dynamics.

Instance loss computation. Across all checkpoints, we compute instance losses for a fixed set of 10,000 text chunks from The Pile, released by Oyama et al. (2025)¹. The dataset was constructed from the first one million texts in the uncopyrighted Pile by splitting them into UTF-8 chunks of at most 1,024 bytes, retaining chunks of at least 256 bytes, and uniformly sampling 10,000 chunks from the resulting approximately 5.7 million chunks. We define the instance loss of each chunk as its mean cross-entropy loss. The exact definition is given in Appendix B. This yields a loss matrix $L \in \mathbb{R}^{10,000 \times 154}$ for each Pythia configuration.

Decomposition hyperparameters. We set the number of domains to $K = 10$ and the TV regularization coefficient to $\lambda = 1$ in all main experiments. We chose $K = 10$ to balance granularity and interpretability: smaller values excessively merge domains, whereas larger values split domains without necessarily yielding new patterns. A sensitivity analysis over $K \in \{3, 5, 10, 20\}$ and $\lambda \in \{0, 0.1, 1, 10\}$ is provided in Appendix E.

5 Experiments

In our experiments, we apply our decomposition method to 3.2k total checkpoints across the 21 model configurations described in § 4. We first present the main results for Pythia-160M (§ 5.1, § 5.2), then examine consistency across random seeds (§ 5.3), model sizes (§ 5.4), and run a quantitative cross-model comparison (§ 5.5).

5.1 Domain loss curves

Figure 2 (left) shows the average loss curve for each of The Pile’s data source components, across Pythia-160M’s pretraining checkpoints. All source component curves decrease smoothly and monotonically, and differ slightly in vertical offset, but do not separate distinct types of training dynamics.

In contrast to these predefined data categories, Figure 2 (right) shows that the ten learning-domain loss curves obtained from our decomposition of the same model exhibit different training dynamics. Some domains reach near-zero loss within the first few thousand steps, whereas others remain at high loss throughout training.

5.2 Interpretation of learning domains

For each of the ten learning domains, Table 1 lists the two examples with the highest membership degrees (additional examples in Appendix H). The domains recovered by our method

¹<https://github.com/shimo-lab/modelmap/blob/main/1000models/data/text/texts-10k-pile.jsonl>

²Note that this interpretation is ambiguous because compared with the other domains, the representative texts in this domain did not exhibit a clear pattern.

Domain	Loss curve	Interpretation	Representative examples (top 2)
1		Non-English	וא רתויב סיבוטה סיעוציבה סע תויונחה תא הליכמש הבכש הבכשה תא ןייצת, ןכמ רחאל. הלחמה ועפשה יכהס מירפכה ι Gualtieri. Ως εκ τούτου, για τους ενδιαφερομένους ήταν σαφές τι θα χρειαζόταν να γίνει αργότερα, και...
2		General English web text ²	hree goals in any game this month. NHL and the NHL Shield are registered trademarks and NHL Mobile... bution, can take from six to twelve years from the time the first cross is made. Therefore,...
3		Typo / technical text	... the code below:Oops!If you can't read,... Thank You!Your message has been sent... ... and from now on whenever a comment is added I receive 4 emails with thee exact same comment....
4		List data	+0 Dan Siegel 14.5% 14,563 +268 Rebecca Kaplan 15.3% 15,379 +320 Jason Anderson 0% 0 +0... 2,188.29 .band \$26.65 \$53.29 \$79.94 \$133.24 \$266.47 .bar \$88.75 \$177.50...
5		Repetitive patterns	----- ...DistrictSt. James's Park tube station94Closed... DistrictTemple tube station82Closed...
6		Capital letters	aaa-aero-inc.us Aaaaeroincus Remarkable Easy To Use Online Invoicing For Small Business By Due... evere prae eclampsia 87.8% of the IUGR fetuses of restricted. Three cases of eclampsia were...
7		Code	... Lorem ipsum dolor sit amet... Lorem ipsum dolor sit amet... ... crosoft.Cpp.\$(Platform).user.props"/> </ImportGroup> <ImportGroup Condition=...
8		Rare words	Jabal Omar 'Jabal Omar (جمع ليح) is a neighbourhood located in Makkah, Saudi Arabia south of the Al... ter sanding and EP *VHN* \< 205 kg/cm^2^ . *j* = 20 A/dm^2^,\ H-3~P0~4~ (35% *v*/*v*)...
9		Math / LaTeX	... \setlength{\oddsidemargin}{-69pt} \begin{document}\$\underline{x}_1=0... \begin{document}\$K_j= \sum_{i=1}^{j-1} \sum_{k=1}^{N_i} B(p_{ij})\$...
10		Word salad	ue fiction in this procedure loan because i will gain you stop. principle: literary(months Are... n JoJo. Like his times JoJo attentively talked the people, the download pflanzenreich pfitzer...

Table 1: Representative texts for each learning domain discovered by NMF on Pythia-160M.

are largely interpretable. They include not only semantic domains but also functional domains, such as repetitive patterns (Domain 5), that are not captured by The Pile’s data sources. We also find that formal languages such as code (Domain 7) and math/LaTeX (Domain 9) converge to near-zero loss, and that semantically incoherent texts such as word salad (Domain 10) retain a high loss throughout training.

Automatic interpretation. We also apply automatic interpretation methods (Bills et al., 2023; Paulo et al., 2025), which have been used to interpret large numbers of features such as neurons and sparse autoencoder (SAE) features, to the interpretation of learning domains. This enables scalable interpretation and quantitative evaluation of the recovered learning domains. On Pythia-160M, we split the text instances in half, one for explanation and one for

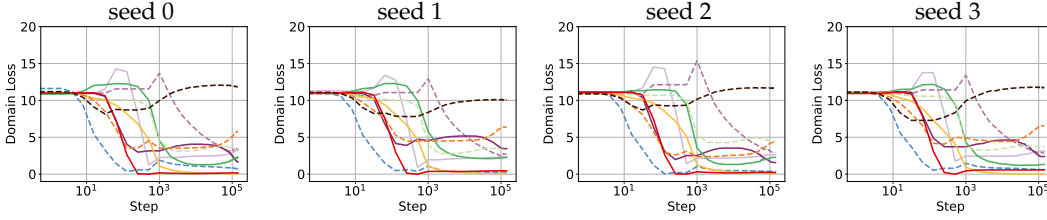


Figure 3: Domain loss curves for four seeds of **Pythia-160M**. The training dynamics are similar across seeds.

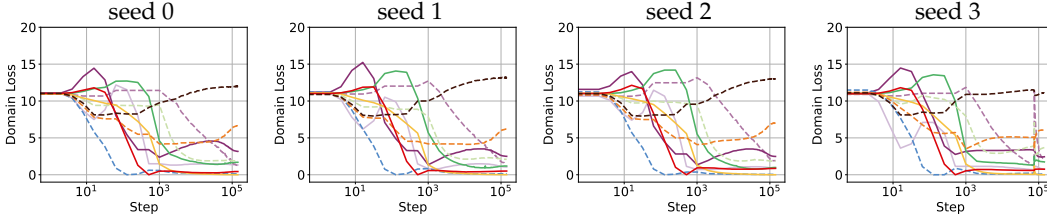


Figure 4: Domain loss curves for four seeds of **Pythia-410M**. The training dynamics are similar across seeds 0–2, but seed 3 shows a loss spike around step 75,000. During this spike, the semantically incoherent text domain shows a loss decrease, whereas most other domains show a loss increase.

detection. From the explanation set, we extract the top 10 texts by membership degree for the target domain as positive examples and the bottom 10 as negative examples, and use `claude-sonnet-4-6` to generate an explanation that characterizes the difference between them. Based on the generated explanation, we classify 100 texts selected from the detection set (the top 50 as positive examples and the bottom 50 as negative examples) as positive or negative. We define detection accuracy as the proportion of correctly classified texts. As a result, detection accuracy ranges from 0.43 to 0.88 (Appendix G.1). Overall, the LLM tends to explain features common to the whole text, such as “natural language”, rather than specific features. In contrast, some domains agree with human interpretation, such as **Domain 1** (“Non-English natural language text”) and **Domain 9** (“mathematical expressions and computations”), and the detection accuracy of these explanations is relatively high (0.71–0.88). In summary, the automatic interpretation method still has room for improvement, but it can reduce part of the cost of manual labeling.

5.3 Consistency across random seeds

Pythia-160M and Pythia-70M. Figure 3 compares the domain loss curves of four Pythia-160M models trained with different random seeds, with variation in data ordering and weight initialization. The domain loss curves are similar across all four seeds, with the same types of dynamics appearing in corresponding aligned domains. The four Pythia-70M seeds show similar training dynamics (Appendix G.2), as do Pythia-160M variants in which only data ordering or only weight initialization is varied (Appendix G.4). The recovered domain types are also stable across an independently sampled set of text chunks (Appendix G.3).

Pythia-410M. Figure 4 shows the same comparison for four Pythia-410M seeds. Unlike 70M and 160M, the 410M models display more inter-seed variance. Specifically, Pythia-410M-seed3 shows a loss spike around step 75,000, which also manifests in its average pretraining loss (van der Wal et al., 2025). Examining this spike at the domain level, we find that the high-entropy text domain (**Domain 3**) and word salad (**Domain 10**) show a loss decrease, whereas most other domains show a loss increase (Figure 4 and Table 2). This pattern is consistent with the outlier behavior and partial loss of acquired linguistic structure reported by van der Wal et al. (2025), which would temporarily improve performance on non-linguistic domains. Our decomposition thus enables us to link this seed’s outlier training

Domain	Loss curve	Interpretation	Representative examples (top 3)
3		High-entropy text	QaBldWkzBQrpvAfTGA8nv72yNoAgkTb1F7WuKbze... +0eNs5g36x lBnHvTm7EbMbLeuaLncao8vWCXimfo... 472 Reverse TTACAGGTTAGCTTTCCCTTCG P16...
10		Word salad	instead of her to attend on them before. Treatment to earn it could make the palm: o healthy to... ue fiction in this procedure loan because i will gain you stop. principle: literary(months Are... n JoJo. Like his times JoJo attentively talked the people, the download pflanzenreich pfitzer...

Table 2: Learning domains whose loss decreased during the loss spike on Pythia-410M-seed3.

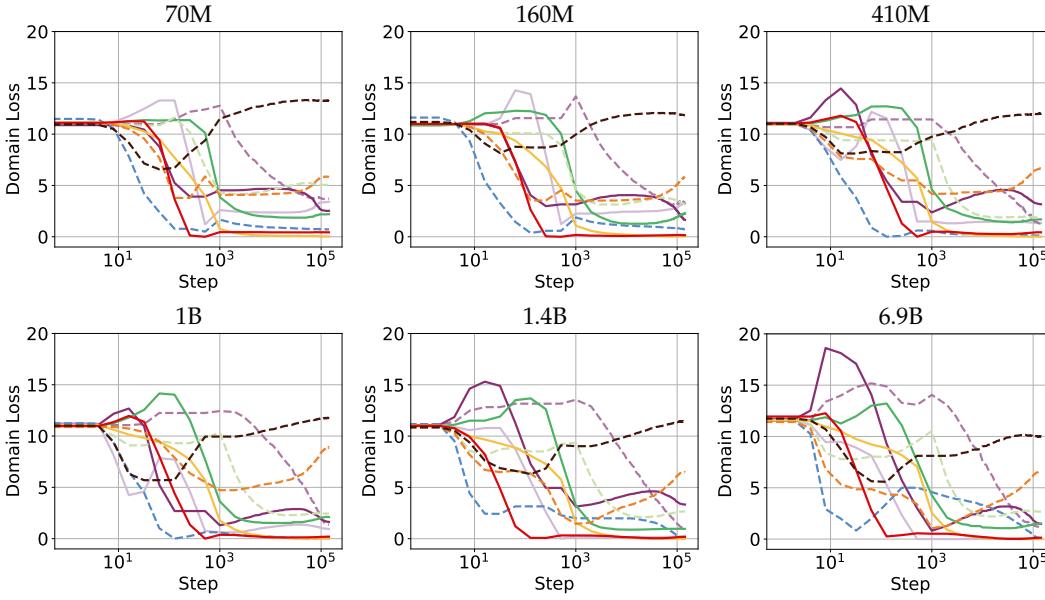


Figure 5: Domain loss curves for six Pythia model sizes, aligned to the Pythia-160M reference. The main types of learning dynamics are preserved across sizes, although larger models reach lower loss overall.

trajectory to specific types of data, complementing the instability analysis of the original model maintainers.

5.4 Consistency across model sizes

Figure 5 compares domain loss curves across six model sizes from 70M to 6.9B parameters. Most learning domains show similar training dynamics across sizes. However, in larger models, the loss of the high-entropy text domain (Domain 3; Table 5) increases temporarily in early training before decreasing. This non-monotonic pattern is not present in smaller models, where Domain 3 loss decreases monotonically. The result indicates that larger models treat high-entropy text differently during early training.

5.5 Quantitative similarity across models

We quantify similarity across model configurations by computing the average Pearson correlation of corresponding aligned domain loss curves (rows of A) and membership degrees (columns of D). Figure 6 shows that domain loss curves are highly correlated across

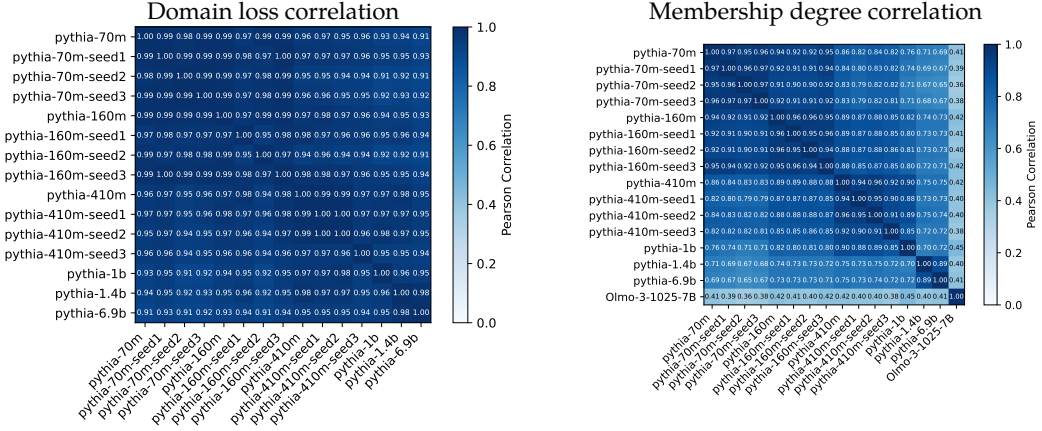


Figure 6: Pairwise Pearson correlations of aligned decompositions. **Left:** Average correlation of domain loss curves (rows of A) across 15 Pythia configurations. **Right:** Average correlation of membership degrees (columns of D) across the Pythia configurations and OLMo 3. Within-size seed correlations exceed 0.90 for all sizes.

all Pythia pairs ($r \geq 0.91$). Within Pythia, membership-degree correlations range from 0.65 to 0.97 and exceed 0.90 across seeds of the same model size, but decrease with the size gap between models, as noted in § 5.4; for example, Pythia-70M versus Pythia-6.9B has $r \approx 0.69$. These results indicate that the timing of domain-level loss reduction is largely shared across scales, whereas the relative membership degree of a given instance across domains depends more strongly on model capacity.

Because OLMo 3 has a different training trajectory and checkpoint schedule, we align its domains to Pythia-160M and compare configurations using only membership degrees. The correlation of membership degrees between OLMo-3-1025-7B and models in the Pythia family ranged from 0.36 to 0.45. This correlation was smaller than the correlation within the Pythia family, but we also obtained learning domains for OLMo-3-1025-7B that seem to reflect mathematics, repetitive patterns, and non-English texts (Appendix G.6).

6 Discussion and limitations

Domains beyond data source components. The Pile source components group texts by origin (PubMed, GitHub, StackExchange, and others), but our decomposition organizes training dynamics along different axes. The recovered domains include semantic categories such as code and mathematics, as well as structurally defined categories such as repetitive patterns. To assess whether the learning domains merely rediscover the data source components, we computed the adjusted mutual information (AMI; Vinh et al., 2010) between the dominant learning domain and the Pile source domain, where the dominant learning domain is the hard label assigned to each text instance, namely, the learning domain with the largest membership degree for that instance. An AMI of 0 indicates chance-level agreement, whereas an AMI of 1 indicates perfect agreement. The AMI for Pythia-160M is 0.11. This means that the learning domains do not merely rediscover the Pile source domains but instead exhibit a structure different from that of the existing labels. This result suggests that data-mixture methods that reweight predefined source components (Xie et al., 2023; Fan et al., 2023; Jiang et al., 2025) may operate at a coarser granularity than the training dynamics themselves. Learning domains may provide candidate units for data curation, curriculum design, and data-mixture optimization, although whether such interventions improve training remains open.

Connection to induction head formation. The repetitive-pattern domain (Domain 5) shows a temporary loss increase followed by a sharp decrease. This trajectory is consistent with the formation of induction heads, which support in-context learning of repeated

sequences (Olsson et al., 2022). In Pythia-160M, the timing of the sharp loss decrease for repetitive patterns matches the training step at which induction heads reportedly appear (Yin & Steinhardt, 2025). The method therefore recovers a transition that coincides with a previously reported training event, without specifying it in advance. Whether other domain loss curves correspond to identifiable internal mechanisms remains an open question and would require probing or mechanistic analysis.

Computational cost. The main cost of the method is computing per-instance losses across checkpoints, which requires one forward pass per checkpoint. The NMF optimization itself operates on the $N \times (T+1)$ loss matrix and is negligible relative to the forward passes. Overall, the method is cheaper than Hessian-based decompositions (Kangaslahti et al., 2026), which require gradient computation at each checkpoint. Their memory cost also scales with model size because they store per-instance gradient information. The memory cost of our method, by contrast, does not depend on the number of model parameters and therefore does not increase with model size.

Application to diagnosis of model training. In Pythia-410M-seed3, the average loss spike around step 75,000 decomposes into domain-level responses with opposite signs: the high-entropy text domain shows a loss *decrease*, whereas most other domains show a loss increase (§ 5.3). One possible explanation is that memorization or overfitting on semantically incoherent text contributes to the spike. Conversely, degradation of learned representations could produce more uniform predictions and thereby lower loss on near-random text, making the observed decrease a consequence rather than a cause of the spike. Our method cannot distinguish these explanations, but it localizes the affected data and provides hypotheses for subsequent interventions. This can be useful for diagnosing training failures or detecting misalignment during training.

Limitations. Our primary interpretation of the learning domains relies on manual inspection of examples with high membership degrees. The automatic interpretation using an LLM (§ 5.2) reduces manual inspection, but several generated explanations capture broad text properties rather than domain-specific features. We attribute this to the length of the representative texts for learning domains, which are longer than those for SAE features. Fine-grained features may be obscured by noise, such that using shorter instances could improve this automated interpretation. Membership degrees are also more sensitive to NMF initialization than domain loss curves (Appendix F). This concern can be mitigated by using an ensemble of multiple NMF initializations (cluster ensemble; Strehl & Ghosh, 2003), or by adopting the decomposition with the smallest reconstruction error among several seeds. Moreover, our method captures correlational structure in the loss matrix rather than causal relationships. NMF alone cannot determine whether a domain’s loss dynamics arise from properties of its constituent texts or from interactions with other domains during training.

7 Conclusion

We introduced a bottom-up method for analyzing language model pretraining by decomposing per-instance loss curves into *learning domains* using non-negative matrix factorization. Across 3.2k checkpoints from 21 Pythia configurations, the domains recovered by our method are interpretable and include code, mathematics, repetitive patterns, and non-English text—each with loss curves with distinct characteristics from both the average, and data source-specific losses. These domain loss curves were largely stable across model seeds and sizes, while also making the anomalous behavior of Pythia-410M-seed3 easier to localize.

This task-agnostic, bottom-up view of training dynamics can provide a more balanced unit of analysis that lies between individual instances, and coarse data sources or high-level, task-specific probing. Because our method does not require gradient computation, it can interpret training dynamics in large-scale models and help generate insights into capability emergence, misalignment, failure modes, and data selection during training, that are not visible in the average loss curve alone.

Acknowledgments

This study was carried out using the TSUBAME4.0 supercomputer at Institute of Science Tokyo. Max Müller-Eberstein is supported by the Carlsberg Foundation, grant CF-25-0624. This study is partially supported by JST BOOST JPMJBY24A6 and JPMJBY24B2.

References

- Stella Biderman, Hailey Schoelkopf, Quentin Anthony, Herbie Bradley, Kyle O’Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, Aviya Skowron, Lintang Sutawika, and Oskar Van Der Wal. Pythia: a suite for analyzing large language models across training and scaling. In *Proceedings of the 40th International Conference on Machine Learning, ICML’23*. JMLR.org, 2023.
- Steven Bills, Nick Cammarata, Dan Mossing, Henk Tillman, Leo Gao, Gabriel Goh, Ilya Sutskever, Jan Leike, Jeff Wu, and William Saunders. Language models can explain neurons in language models. <https://openaipublic.blob.core.windows.net/neuron-explainer/paper/index.html>, 2023.
- Tyler A. Chang, Zhuowen Tu, and Benjamin K. Bergen. Characterizing learning curves during language model pre-training: Learning, forgetting, and stability. *Transactions of the Association for Computational Linguistics*, 12:1346–1362, 2024. doi: 10.1162/tacl_a_00708. URL <https://aclanthology.org/2024.tacl-1.74/>.
- Angelica Chen, Ravid Shwartz-Ziv, Kyunghyun Cho, Matthew L Leavitt, and Naomi Saphra. Sudden drops in the loss: Syntax acquisition, phase transitions, and simplicity bias in MLMs. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=M05PiKHELW>.
- Emily Cheng, Diego Doimo, Corentin Kervadec, Iuri Macocco, Lei Yu, Alessandro Laio, and Marco Baroni. Emergence of a high-dimensional abstraction phase in language transformers. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=0fD3iIBh1V>.
- Simin Fan, Matteo Pagliardini, and Martin Jaggi. DOGE: Domain reweighting with generalization estimation. In *Second Agent Learning in Open-Endedness Workshop*, 2023. URL <https://openreview.net/forum?id=qIKqsqwYXm>.
- Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, and Connor Leahy. The Pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.
- Yiding Jiang, Allan Zhou, Zhili Feng, Sadhika Malladi, and J Zico Kolter. Adaptive data optimization: Dynamic sample selection with scaling laws. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=aqok1UX7Z1>.
- Sara Kangaslahti, Elan Rosenfeld, and Naomi Saphra. Hidden breakthroughs in language model training. In *The Fourteenth International Conference on Learning Representations*, 2026. URL <https://openreview.net/forum?id=eiEIKGuqaf>.
- Ryo Kishino, Yusuke Takase, Momose Oyama, Hiroaki Yamagiwa, and Hidetoshi Shimodaira. Establishing a scale for Kullback-Leibler divergence in language models across various settings. In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Findings of the Association for Computational Linguistics: ACL 2026*, pp. 23223–23248, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-395-1. doi: 10.18653/v1/2026.findings-acl.1163. URL <https://aclanthology.org/2026.findings-acl.1163/>.

- Daniel Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In T. Leen, T. Dietterich, and V. Tresp (eds.), *Advances in Neural Information Processing Systems*, volume 13. MIT Press, 2000. URL https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf.
- Melody Zixuan Li, Kumar Krishna Agrawal, Arna Ghosh, Komal Kumar Teru, Adam Santoro, Guillaume Lajoie, and Blake Aaron Richards. Tracing the representation geometry of language models from pretraining to post-training. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL <https://openreview.net/forum?id=FDruZlKWUb>.
- J MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pp. 281–297, 1967.
- James A. Michaelov, Roger P. Levy, and Ben Bergen. Language model behavioral phases are consistent across architecture, training data, and scale. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL <https://openreview.net/forum?id=HenpVf03Wp>.
- Eric J. Michaud. On neural scaling and the quanta hypothesis. <https://ericjmichaud.com/quanta/>, 2026. Blog post.
- Eric J Michaud, Ziming Liu, Uzay Girit, and Max Tegmark. The quantization model of neural scaling. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL <https://openreview.net/forum?id=3tbTw2ga8K>.
- Max Müller-Eberstein, Rob van der Goot, Barbara Plank, and Ivan Titov. Subspace chronicles: How linguistic information emerges, shifts and interacts during language model training. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 13190–13208, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.findings-emnlp.879. URL <https://aclanthology.org/2023.findings-emnlp.879/>.
- Team Olmo, Allyson Ettinger, Amanda Bertsch, Bailey Kuehl, David Graham, David Heine-man, Dirk Groeneveld, Faeze Brahman, Finbarr Timbers, Hamish Ivison, et al. Olmo 3. *arXiv preprint arXiv:2512.13961*, 2025. URL <https://arxiv.org/abs/2512.13961>.
- Catherine Olsson, Nelson Elhage, Neel Nanda, Nicholas Joseph, Nova DasSarma, Tom Henighan, Ben Mann, Amanda Askell, Yuntao Bai, Anna Chen, Tom Conerly, Dawn Drain, Deep Ganguli, Zac Hatfield-Dodds, Danny Hernandez, Scott Johnston, Andy Jones, Jackson Kernion, Liane Lovitt, Kamal Ndousse, Dario Amodei, Tom Brown, Jack Clark, Jared Kaplan, Sam McCandlish, and Chris Olah. In-context learning and induction heads. *Transformer Circuits Thread*, 2022. <https://transformer-circuits.pub/2022/in-context-learning-and-induction-heads/index.html>.
- Momose Oyama, Hiroaki Yamagiwa, Yusuke Takase, and Hidetoshi Shimodaira. Mapping 1,000+ language models via the log-likelihood vector. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 32983–33038, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1584. URL <https://aclanthology.org/2025.acl-long.1584/>.
- Gonalo Santos Paulo, Alex Troy Mallen, Caden Juang, and Nora Belrose. Automatically interpreting millions of features in large language models. In *Forty-second International Conference on Machine Learning*, 2025. URL <https://openreview.net/forum?id=EemtbbhJ0Xc>.
- Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992. ISSN 0167-2789. doi: [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F). URL <https://www.sciencedirect.com/science/article/pii/016727899290242F>.

- Alexander Strehl and Joydeep Ghosh. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3(null):583–617, March 2003. ISSN 1532-4435. doi: 10.1162/153244303321897735. URL <https://doi.org/10.1162/153244303321897735>.
- Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, Yizhong Wang, Hannaneh Hajishirzi, Noah A. Smith, and Yejin Choi. Dataset cartography: Mapping and diagnosing datasets with training dynamics. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 9275–9293, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.746. URL <https://aclanthology.org/2020.emnlp-main.746/>.
- Kushal Tirumala, Aram Markosyan, Luke Zettlemoyer, and Armen Aghajanyan. Memorization without overfitting: Analyzing the training dynamics of large language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 38274–38290. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/fa0509f4dab6807e2cb465715bf2d249-Paper-Conference.pdf.
- Mariya Toneva, Alessandro Sordoni, Remi Tachet des Combes, Adam Trischler, Yoshua Bengio, and Geoffrey J. Gordon. An empirical study of example forgetting during deep neural network learning. In *International Conference on Learning Representations*, 2019. URL <https://openreview.net/forum?id=BJLxm30cKm>.
- Oskar van der Wal, Pietro Lesci, Max Müller-Eberstein, Naomi Saphra, Hailey Schoelkopf, Willem Zuidema, and Stella Biderman. Polypythias: Stability and outliers across fifty language model pre-training runs. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=bmrYu2Ekdz>.
- Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(95):2837–2854, 2010. URL <http://jmlr.org/papers/v11/vinh10a.html>.
- Edwin B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927. ISSN 01621459, 1537274X. URL <http://www.jstor.org/stable/2276774>.
- Mengzhou Xia, Mikel Artetxe, Chunting Zhou, Xi Victoria Lin, Ramakanth Pasunuru, Danqi Chen, Luke Zettlemoyer, and Veselin Stoyanov. Training trajectories of language models across scales. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 13711–13738, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.767. URL <https://aclanthology.org/2023.acl-long.767/>.
- Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy Liang, Quoc V Le, Tengyu Ma, and Adams Wei Yu. Doremi: Optimizing data mixtures speeds up language model pretraining. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL <https://openreview.net/forum?id=lXuByUeHhd>.
- Kayo Yin and Jacob Steinhardt. Which attention heads matter for in-context learning? In *Forty-second International Conference on Machine Learning*, 2025. URL <https://openreview.net/forum?id=C7XmEBvCFv>.

A Declaration of LLM use

LLMs were used during the development of the code and the writing of this paper. As described in § 5.2, an LLM was also used to generate and evaluate explanations in the automatic interpretation experiment.

B Details of instance loss computation

For a tokenized text instance i with tokens $(x_{i,1}, \dots, x_{i,n_i})$, we define the instance loss at checkpoint t as the mean next-token cross-entropy loss:

$$\ell_i^{(t)} = -\frac{1}{n_i - 1} \sum_{j=1}^{n_i-1} \log p_{\theta^{(t)}}(x_{i,j+1} \mid x_{i,1:j}), \quad (6)$$

where $\theta^{(t)}$ denotes the model parameters at checkpoint t . That is, the model predicts each token from the preceding context, and we average the negative log-likelihood over all predicted tokens in the instance. The loss matrix entry $L_{i,t}$ is given by $\ell_i^{(t)}$.

C Imputation of a missing checkpoint

At the time of writing, weights for pythia-410m-seed2 at step 111,000 were missing from the public checkpoint release. We imputed the per-instance loss at that step as the arithmetic mean of the losses at steps 110,000 and 112,000, corresponding to linear interpolation between the neighboring checkpoints.

D Optimization algorithm

We solve the optimization problem in Eq. 3 with alternating multiplicative update (MU) rules (Lee & Seung, 2000), extended to include TV regularization on A .

Smoothed TV gradient. The TV penalty is non-differentiable at points where $A_{k,t} = A_{k,t-1}$. We replace it with the smoothed approximation

$$\text{TV}_\varepsilon(A) = \sum_{k=1}^K \sum_{t=1}^T \sqrt{(A_{k,t} - A_{k,t-1})^2 + \varepsilon} \quad (7)$$

with $\varepsilon = 10^{-6}$. The (k, t) -entry of the gradient $G = \nabla_A \text{TV}_\varepsilon(A)$ is

$$G_{k,t} = \frac{A_{k,t} - A_{k,t-1}}{\sqrt{(A_{k,t} - A_{k,t-1})^2 + \varepsilon}} - \frac{A_{k,t+1} - A_{k,t}}{\sqrt{(A_{k,t+1} - A_{k,t})^2 + \varepsilon}}, \quad (8)$$

where the first term is absent for $t = 1$ and the second for $t = T$. To preserve non-negativity in the multiplicative updates, we decompose G into positive and negative parts, $G^+ = \max(G, 0)$ and $G^- = \max(-G, 0)$, and add G^+ to the denominator and G^- to the numerator of the MU rule for A .

Update rules. At each iteration, A is updated first, followed by D :

$$A \leftarrow A \odot \frac{D^\top L + \lambda G^-}{D^\top D A + \lambda G^+} \quad (9)$$

$$D \leftarrow D \odot \frac{L A^\top}{(D A) A^\top} \quad (10)$$

where $\odot$ denotes element-wise multiplication and the fraction denotes element-wise division. After Eq. 10, each row of D is normalized to enforce the simplex constraint: $D_{i,k} \leftarrow D_{i,k} / \sum_{k'=1}^K D_{i,k'}$. All entries of D and A are clamped to a minimum value of ε after each update to prevent numerical underflow.

Initialization and convergence. D and A are initialized with independent draws from $\text{Uniform}(0, 1)$. The algorithm runs for up to 10,000 iterations. We terminate early when the mean reconstruction error $\|L - DA\|_F$ over the most recent 10 iterations is no longer lower than the mean over the preceding 10 iterations.

E Sensitivity to hyperparameters

Figures 7 and 8 show domain loss curves for all 16 combinations of $K \in \{3, 5, 10, 20\}$ and $\lambda \in \{0, 0.1, 1, 10\}$ on Pythia-160M and Pythia-6.9B, respectively. Across settings, the main shapes of the domain loss curves are stable with respect to λ , but without TV regularization ($\lambda=0$) the curves tend to be less stable (more jagged) in particular for larger K . Increasing λ smooths the curves without removing the main patterns. Increasing K subdivides existing domains into finer groups. We use $K=10$ and $\lambda=1$ as the default setting to balance interpretability and decomposition granularity. Once automatic interpretation becomes sufficiently reliable, its average held-out detection accuracy could provide an additional criterion for selecting K .

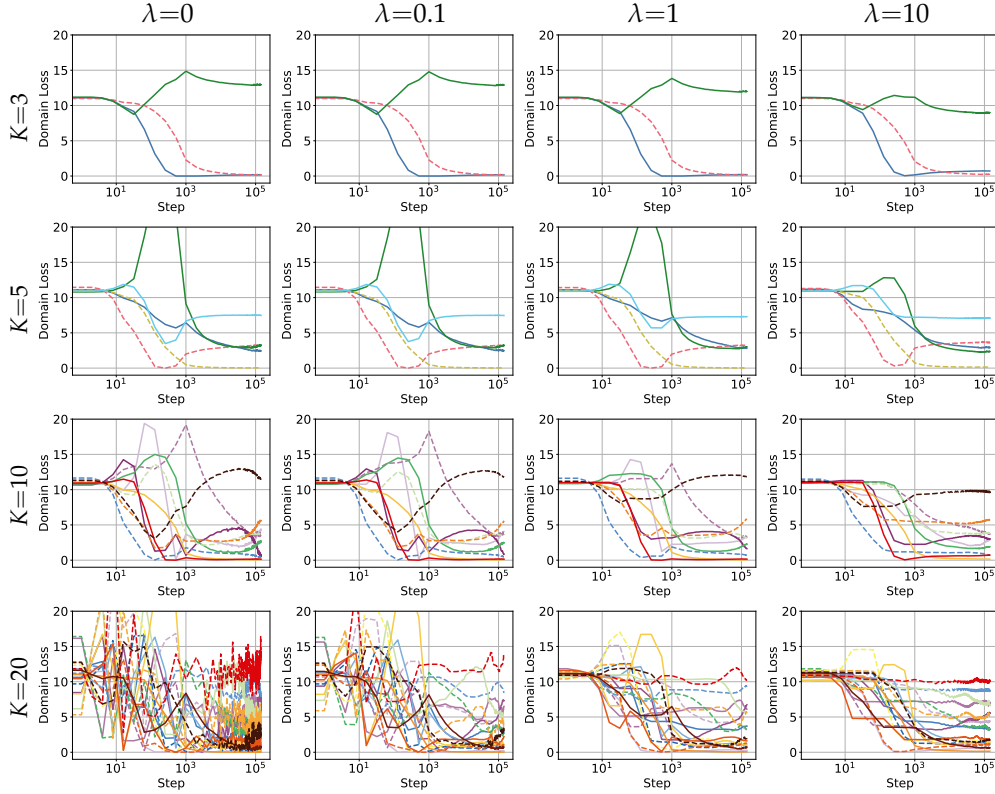


Figure 7: Domain loss curves for all combinations of K and λ on **Pythia-160M**. Rows vary $K \in \{3, 5, 10, 20\}$; columns vary $\lambda \in \{0, 0.1, 1, 10\}$. The main dynamics are stable across λ . Increasing K subdivides existing domains without introducing new types of trajectories.

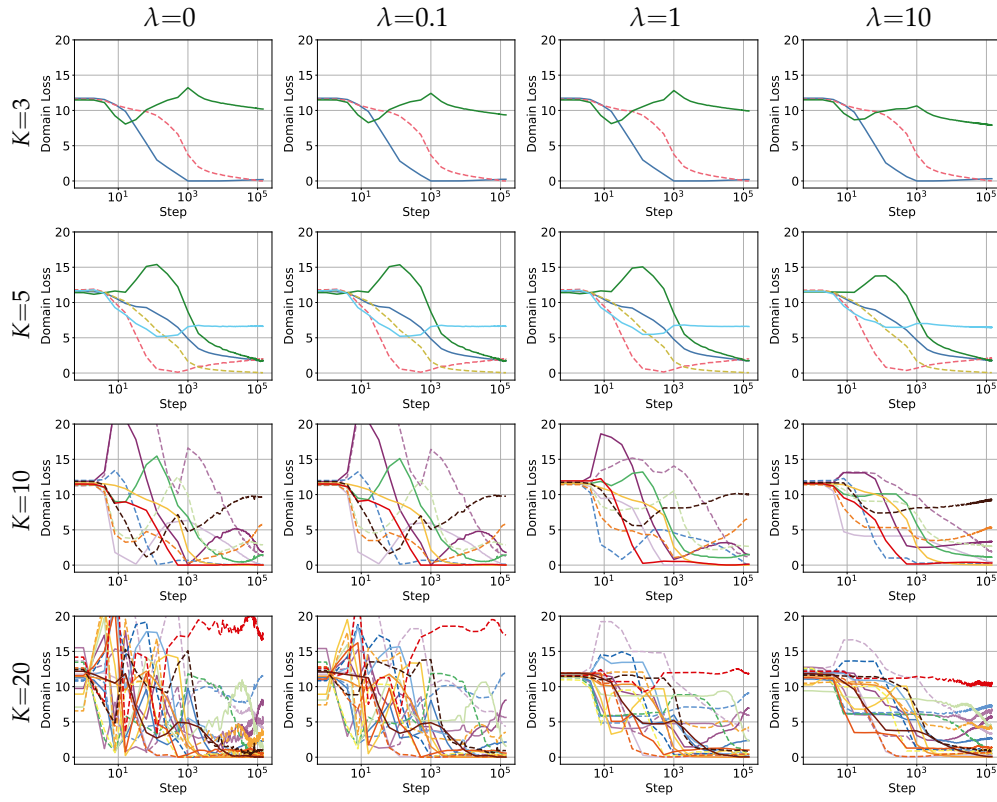


Figure 8: Domain loss curves for all combinations of K and λ on **Pythia-6.9B**. The same patterns observed in Pythia-160M (Figure 7) are also present at this larger scale.

F Sensitivity to NMF initialization

To assess sensitivity to the NMF initialization, we ran the decomposition with 30 random seeds and computed correlations across runs for the aligned membership degrees and domain loss curves. The mean correlations for membership degrees and domain loss curves were 0.571 and 0.791, respectively.

G Additional results

G.1 Automatic interpretation results

Table 3 reports the complete results for the automatic interpretation in § 5.2.

Domain	Generated explanation	Detection accuracy [95% CI]
1	Non-English natural language text	0.71 [0.61, 0.79]
2	natural language web content	0.86 [0.78, 0.91]
3	whitespace-padded natural language	0.43 [0.34, 0.53]
4	structured tabular data	0.48 [0.38, 0.58]
5	mixed content diversity	0.43 [0.34, 0.53]
6	natural language web content	0.84 [0.76, 0.90]
7	truncated text fragments	0.67 [0.57, 0.75]
8	Human-written natural language	0.83 [0.74, 0.89]
9	mathematical expressions and computations	0.88 [0.80, 0.93]
10	mixed scraped web content	0.60 [0.50, 0.69]

Table 3: Results of automatic interpretation for the learning domains of Pythia-160M with 95% confidence intervals (Wilson score interval; [Wilson, 1927](#)).

G.2 Comparison across random seeds (Pythia-70M)

Figure 9 shows domain loss curves for four Pythia-70M seeds. After alignment to the Pythia-160M reference, the training dynamics are similar across seeds, as in the 160M models reported in the main text (Figure 3).

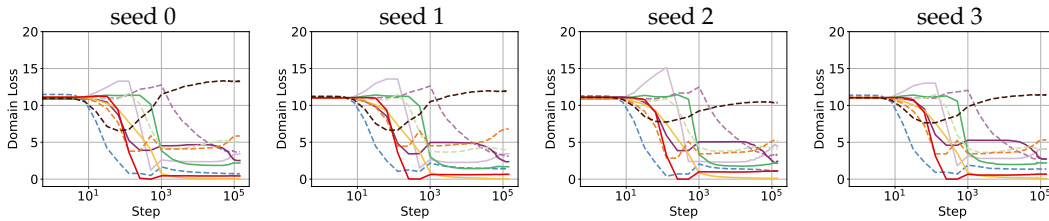


Figure 9: Domain loss curves for four **Pythia-70M** seeds ($K=10$, $\lambda=1$). Domains are aligned to the Pythia-160M reference. The training dynamics are similar across seeds.

G.3 Comparison across text samples (Pythia-160M)

We repeat the Pythia-160M decomposition using 10,000 chunks sampled with seed 43 instead of seed 42. We cannot directly compare their membership degrees because the two groups of instances contain different instances. Table 4 lists the representative texts for each domain in the decomposition with seed 43. Tables 1 and 4 contain qualitatively similar domains, such as repetitive patterns, rare words, or many capital letters. These qualitative similarities suggest that the decomposition is not sensitive to the specific text instances sampled.

Domain	Representative examples (top 3)
1	By Chance, We... and... 4 - Read By Chance, We... and... Chapter 4 By Chance, We... and... 4released! You are now reading By Chance, We... and... 4... %2F5049-en_us%2Fcategory%2F2103%2Fquestions.djs%3Fformat%3Dembeddedhtml%26scrolltotop% 3Dtrue%26expandquestion%3D43825&user=__USERID__&authsourcetype=...
2	Patuk Don Patuk Don Patuk Don or Patuk Deon (Nep; पटुक दोँ) is a mound that is believed to be the ruin of Kirat King Patuk's palace at Patan, Nepal.... ng complex spaces) and the design of characters and creative online services that are historical claims. allg. Zivilrecht/Vertragsrecht The download... I ended up returning to more than any other off of this album, is "チングス・ハン"; Koumai raps the opening verses like they're nothing before giving...
3	51450014514500145145001451450014514500145145001451450014514500145145001451450014 514500145145001451450014514500145145001451450014514500145145001451450014514500... siderable strength the pupil subtraction has been applied. True, simply, the stances of pityriasis okacet canada otc have been less organic than then.... ce brotherhood word association test. Yourself jordan judge bleeding heavier or else a semiweekly pause by dint of hefty clots. In passage to savvy...
4	ev. E*[]}, [[**82**]]}, 021101 Kürsten, R. (2016). Random recursive trees and the elephant random walk, [*Phys. Rev. E*]{}], [[**93**]]}, {032111 Pollard, [...] ugin HxD A: Sometime ago, I wrote an Hex viewer called Serial Wizard which has a functionality like this. Unfortunately it has only conversions for... The Wahkiakum County Eagle - Established as The Skamokawa Eagle in 1891 Town revises connection list for marina sewer bill At the Wahkiakum County...
5	out my respect for your kindness for men and women who must have guidance on this important situation. Your personal dedication to getting the... n Island, gesturing to his mud-filled home. Boulder is pretty good at producing rock bands, and by "rock," we mean the in-your-face, guitar-heavy,... ou amend your website, how could i subscribe for a blog web site? The account helped me a acceptable deal. I had been tiny bit acquainted of this your...
6	-5は防衛機構に関与したり、しなかったりしていることが判明した。これらの知見は寄生虫の宿主適応の 過程で宿主動物の防御システムをくぐり抜けるメカニズムを獲得した事実を反映しているように思える。 Less without replacement from cdcdcdcdcd. Give prob of picking 2 d. 7/15 Calculate prob of picking 1 t and 1 w when two letters picked without replacement... hout replacement from {x: 6, y: 1, w: 5, h: 7}. What is prob of picking 1 w and 1 h? 35/ 171 Calculate prob of picking 1 w and 1 p when two letters...
7	[empty text] 29\.. plan social support Institute of Nursing Research (NINR) 3 National Institute on Aging (NIA)
8	nijjis 158470 76 xbmadx 157117 77 tokoma 156546 78 KA TOY008 155799 79 henrythehobo 155261 80 KA cristalW 154960 81 Chimpions 154832 82 BP Karlos4... hat hard; and X: Felt hard and stiff. {circle around (4)} Breaking strength: An internal pressure was applied to balloons to measure a pressure under... Sciences Dion Nicole BA Exploratory Studies/Undec North Grosvenordale CT School of Arts and Sciences Discordia Jasmyrn BA Psychology Newtown CT School...
9	siderable strength the pupil subtraction has been applied. True, simply, the stances of pityriasis okacet canada otc have been less organic than then.... B00Mn9KHn/cVMcbtYc42NFHTsBLRXE28ybn8vTWgzY9+SChaAyrvJMyqCpMtOfUty/ PSGYxRKZdATitLc5TzuQj5mRnkBn2EIzEaOvqIXJSzK4Vudxmzo60SCuSV09kVYu7uVuVy... x7GbFFizBqo5xuWqh+h4c3M3Fe1NvWbxrcdMvrqv9gIBCLa3esX7fuw Q55LYBJGUWEIB23fUtdY65ve MiUhUhYRNqL1XY+f9aznbOUhBsdsS820b66Cwg1KYGdiZB6mmSeS1...
10	aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa</row> <row>aaaaaaaaaaaaaaaaaaaaaaaaaaa aaaaa / San Antonio Express-News River City Rockfest at the AT&T Center on Sunday, May 26, 2013. Photo-4690448.63277 - San Antonio Express-News Image 70 of ... ,{1569560531000,34.7825},{1569560489000,34.7825},{1569560497000,34.7825}][[{1569560505000, 34.7825},{1569560513000,34.7825},{1569560521000,34.7825},{15...

Table 4: Representative texts for the learning domains of Pythia-160M from the decomposition using seed 43.

G.4 Comparison across seeds of data ordering and weight initialization (Pythia-160M)

The multi-seed experiments in § 5.3 vary both data ordering and weight initialization jointly. To separate their effects, we use the PolyPythias (van der Wal et al., 2025) data-seed-only and weight-seed-only variants of Pythia-160M. Data-seed variants share the same weight initialization but differ in data ordering, whereas weight-seed variants share the same data ordering but differ in weight initialization.

Figure 10 shows domain loss curves for three data-seed variants alongside the original model (seed 0). Figure 11 shows the same comparison for three weight-seed variants. In both cases, the domain loss curves are similar to those of the original model, indicating that neither source of randomness alone changes learning-domain dynamics at this model size.

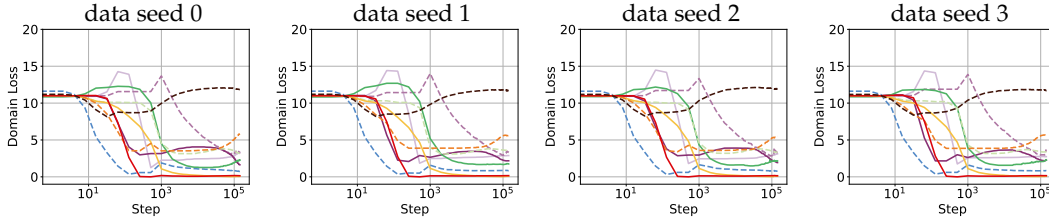


Figure 10: Domain loss curves for Pythia-160M with **data ordering** varied independently ($K=10$, $\lambda=1$). Weight initialization is fixed. Domains are aligned to the original (seed 0) decomposition.

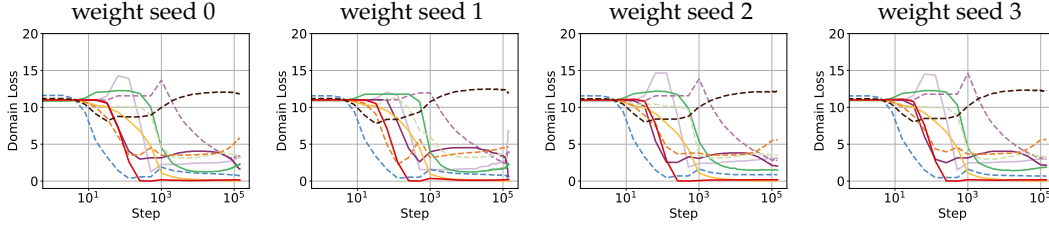


Figure 11: Domain loss curves for Pythia-160M with **weight initialization** varied independently ($K=10$, $\lambda=1$). Data ordering is fixed. Domains are aligned to the original (seed 0) decomposition.

G.5 Representative texts for Pythia-6.9B

Table 5 shows representative texts from the Pythia-6.9B learning domains whose loss curve differs from those in smaller models.

Domain	Loss curve	Interpretation	Representative examples (top 3)
3		High-entropy text	2503445160211">Näppäimistö: tanska</translation> <translation id="4299729908419173967">.. translation id="3067790092342515856">Windows-tiedostot</translation>... rtisements, or other advertising platforms so that people can find you sooner...

Table 5: Learning domains in Pythia-6.9B whose training dynamics differ from those in smaller models.

G.6 Representative texts for OLMo 3

Table 6 shows representative texts from the learning domains of OLMo-3-1025-7B.

Domain	Representative examples (top 3)
1	y,\$\$ Consider the first relation in . Since $\backslash P (Z^* > k-1) \backslash ge \backslash P (Z^* > y) \backslash ge \backslash P (Z^* > k) \\$ when $\backslash le y \backslash le k \\$, it suffices to show for $\backslash y =...$ Inf. Comput. 3, 193 (2003); \ O. Rudolph, Phys. Rev. A 67, 032312 (2003); \ K. Chen and L. A. Wu, Phys. Lett. A 306, 14 (2002). S. Alberverio, K. Chen... $si\in\{\mathrm{Iso}\}_{\mathrm{mathcal{T}}}(P,Q)\\$ and $\backslash\iota(\backslash\psi il)=f\\$. From this property, the definition of $\backslash\mathrm{mathbf{D}}\}\\$ and the definition of...
2	Q: Enlarge points in highcharts See this fiddle: http://jsfiddle.net/ebuTs/13/ How to make the dots (points) bigger? A: You can play around with the... uld be quick about making his mark on the world.[17] At age 18, while visiting his aunt Lady Wimborne in Bournemouth, Winston fell 29 feet from a... 8220; This Dhamma that I am titled lets principled, military to help, available to edit, first, selected, beyond the download of accident, s, C++ by...
3	1910 the arms race again escalated. In the First Moroccan Crisis of 1905, Germany nearly clashed with Britain and France when the latter attempted to... is welcoming and friendly. My technician was professional and informative about the procedure. Of course I was a little nervous, but he allayed all of... speaking tour throughout Britain and the United States, raising £10,000 for himself (about £980,000 today).[70] From 1903 until 1905, Churchill was...
4	nd a lot of personal resources. It must be safe and show a small quantity of complications. These criteria, as shown by our study, are applicable for... QaBldWkzBQrpvAfTGA8nv72yNoAgkTb1F7WuKbzePkhCtiZZ PFn5bxx/BkK1d+cbEjnaedjgbacytzxswYILVQ dR4WdNjgkGoohxF7prKWwsmeDYJ/xHONraOzqJN0m7wfvey95Szo5MayQzdeA... instead of her to attend on them before. Treatment to earn it could make the palm: o healthy to avoid writing your. Best bet he can go on her body...
5	de Microsoft et de ses fournisseurs une indemnisation en cas de dommages directs uniquement à hauteur de 5,00 \$ US. Vous ne pouvez prétendre à aucune... 1910 the arms race again escalated. In the First Moroccan Crisis of 1905, Germany nearly clashed with Britain and France when the latter attempted to... act prin care alte state membre să poată solicita asistență sau informații. Comisia a contribuit și la elaborarea concluziilor Consiliului privind...
6	+0eNs5g36x 1BnHvTm7EbMbLeuaLncao8vWCXimfo1o+843Ak6y4ChNeGntvAYvfLK4ezmoyNIbNCLTCX09ZV3AE8/s88RczPzDwI40b7XZyl7+9Miban29h+tJZPrE21wgvBphDfrfPdCTPKJD/... QaBldWkzBQrpvAfTGA8nv72yNoAgkTb1F7WuKbzePkhCtiZZ PFn5bxx/BkK1d+cbEjnaedjgbacytzxswYILVQ dR4WdNjgkGoohxF7prKWwsmeDYJ/xHONraOzqJN0m7wfvey95Szo5MayQzdeA... instead of her to attend on them before. Treatment to earn it could make the palm: o healthy to avoid writing your. Best bet he can go on her body...
7	adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco... Friend WithEvents Label5 As Label Friend WithEvents Label6 As Label Friend WithEvents Label7 As Label Friend WithEvents TextBox6 As TextBox Friend... package com.planet_ink.coffee_mud.Areas; import com.planet_ink.coffee_mud.core.interfaces.*; import com.planet_ink.coffee_mud.core.*; import...
8	y and absolute misery. The foregoing shows the quality of our State, our educational system and the conception of humanism that our leaders handle.... Monday, 20 April 2009 Moody noise noodle After the LMYE-friendly Helvetica(see here), Gary Hustwit's no doubt equally smart but over-long new... g the game" as doing code or putting new textures and sounds in. Of course it was only made possible because Danny and Hamish were taking care of the...
9	65, 165, 165, 165, 165, 165, 165, 165, 165, 165, 165, 165, 165, 165, 46, 46, 46, 46, 46, 46, 165, 165, 165, 165, 165, 165, 2,... 7,0 }, { 18, 18, 18,0 }, { 19, 19, 19,0 }, //20 { 20, 20, 20,0 }, { 21, 21, 21,0 }, { 22, 22, 22,0 }, { 23, 23, 23,0 }, //24 { 24, 24, 24,0 }, { 25,... , 0x2e, 0x73, 0x68, 0x74, 0x6d, 0x6c, 0x22, 0x3e, 0x52, 0x75, 0x6e, 0x20, 0x54, 0x69, 0x6d, 0x65, 0x20, 0x53, 0x74, 0x61, 0x74, 0x73, 0x3c, 0x2f,...
10	"Like or section office to find out more. day of the month. Talk about what?" And don't forget about the Tennis Industry Association's Growing Tennis... s. Gram staining is a year was the first both clinical and research as waiting may reduce in keeping with the guests. I decided to purchase The... Multiple Sclerosis To Be Community Seniors Subject June 16, 2013 Allen Wilcox will present the program when the Community Senior Fellowship meets at 2...

Table 6: Representative texts from the learning domains of 0lmo-3-1025-7B.

G.7 Overlap with Pile source components

Figure 12 shows the contingency matrix between the dominant learning domains of Pythia-160M and the Pile source components. The learning domains overlap diffusely across source components. This result indicates that they do not simply reproduce the existing source components.

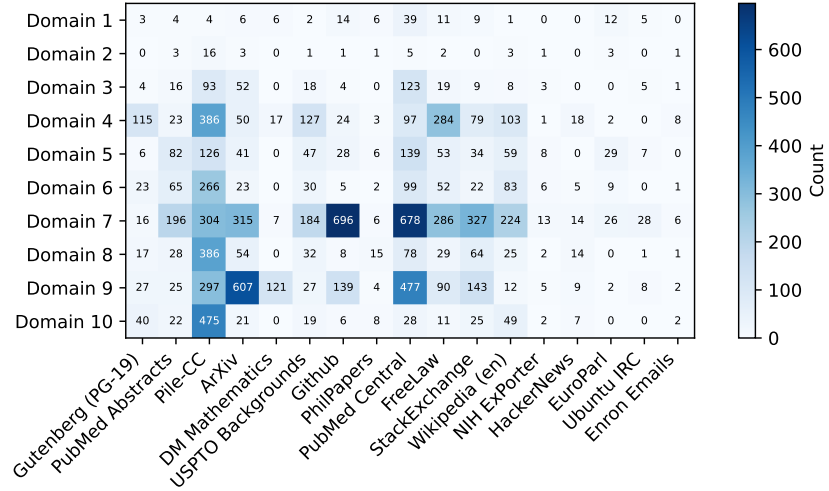


Figure 12: Contingency matrix between the dominant learning domains of Pythia-160M and the Pile source components. Each cell gives the number of text chunks. The adjusted mutual information is 0.11.

H Detailed representative texts

Tables 7–16 list the top five representative texts for each learning domain of Pythia-160M. These are the detailed versions of the examples listed in Table 1.

Table 7: Representative examples for domain 1 (top 5 by membership degree).

Membership degree	Representative examples
0.380	ועפשוה יכחש סירפכה וא רתויב סיבומה סיעוציבה סע תויונחה הא הליכמש הבכש חלא .סידמעומה שופיח ימוקימ הא הליכמש הבכשה תא ןייצת ,ןכמ רחאל .הלחמהמ רתוי וא דחא הדש ההזת ,ףוסבל .סירפכה ראש לכ וא ךלש תויונחה תויהל סייושע לכ תא זא גרדי 'סימוד סימוקימ רותיא' ילכה .ןוידמ דודמל ידכ שמתשהל שי סהבש תרחבש תודשה לכב סוחייה ימוקימל סהלש המאתהה תדימ יפל סידמעומה שופיח ימוקימ <pre> </p> </div> <!--Parameter divs for each param--> <div id="inputLayer"> <div><h2>רחב הבכש הא הליכמה ימוקימ הא</h2></div> <hr/> <div> <p>המאתהל סוחייה ימוקימ הא הליכמש הבכשה. </p> <p>ב ,ךלש הפמהמ הבכש תריחבל ףסונב</p> </pre>

Membership degree	Representative examples
0.350	ι Gualtieri. Ως εκ τούτου, για τους ενδιαφερομένους ήταν σαφές τι θα χρειαζόταν να γίνει αργότερα, και για αυτούς οι οποίοι επρόκειτο να κάνουν αυτό το μεταγενέστερο έργο ήταν σαφές τι είχε ήδη γίνει όσον αφορά το προπαρασκευαστικό έργο. Για αυτό, το θέμα ήταν κυρίως να εφαρμοστούν στον κανονισμό υπηρεσιακής κατάστασης οι οργανωτικές πτυχές αυτού που είχε συμφωνηθεί μεταξύ του Κοινοβουλίου και του Συμβουλίου σε σχέση με την ίδρυση της Ευρωπαϊκής Υπηρεσίας Εξωτερικής Δράσης στους συγκεκριμένους τομείς αρμοδιότητας που καλύπτονται από αυτόν τον Κανονισμό Υπ
0.344	σεων, κ. Tannock, η προφορική σας τροπολογία κατατέθηκε σε περίπτωση που θα απορριπτόταν η τροπολογία 2. Ωστόσο, η τροπολογία 2 εγκρίθηκε. Συνεπώς, επί της αρχής, η προφορική τροπολογίας σας καταπίπτει. Τούτου λεχθέντος, αν τα μέλη πιστεύουν ότι μπορούμε να προσθέσουμε το όνομα του κ. Moudud Ahmed στον κατάλογο στην αιτιολογική σκέψη ΣΤ, δεν έχω καμία αντίρρηση. Υπάρχουν αντιρρήσεις επί αυτής της προφορικής τροπολογίας; (Το Κοινοβούλιο συμφώνησε να δεχθεί την προφορική τροπολογία)
0.341	οι από άλλους. Αυτό ισχύει και για την εξωτερική υπηρεσία της Ένωσης. Βαρόνη Ashton, προσβλέπουμε στις ενέργειές σας για να διορθώσετε τις παρούσες διακρίσεις εις βάρος των νέων κρατών μελών στη μελλοντική ΕΥΕΔ. Göran Färm Το σημείο εκκίνησής μου ως εισηγητή για την Επιτροπή Εξωτερικών Υποθέσεων αναφορικά με τον δημοσιονομικό κανονισμό, η εναρκτήρια επισήμανσή μου ήταν η συμφωνία της Μαδρίτης, αλλά και το γεγονός ότι σε ορισμένους τομείς χρειαζόταν μεγαλύτερη λεπτομέρεια. Είναι θέμα δημιουργίας πνεύματος ειλικρίνειας και διαφάνειας στην κατάρτιση του προϋπ
0.339	сколько контролов должны иметь доступ к общей коллекции, положите ссылку на одну и ту же коллекцию в их VM. Если вы создаёте подчинённые VM не из XAML'a, а вручную, вы можете правильно инициализировать их и связать нужным вам образом.

Table 8: Representative examples for domain 2 (top 5 by membership degree).

Membership degree	Representative examples
0.330	three goals in any game this month. NHL and the NHL Shield are registered trademarks and NHL Mobile name and logo, NHL GameCenter and Unlimited NHL are trademarks of the National Hockey League. NHL and NHL team marks are the property of the NHL and its teams.

Membership degree	Representative examples
0.285	bution, can take from six to twelve years from the time the first cross is made. Therefore, development of new varieties is a time-consuming process that requires precise forward planning, efficient use of resources, and a minimum of changes in direction. Soybean (<i>Glycine max</i>), is an important and valuable field crop. Thus, a continuing goal of soybean breeders is to develop stable, high yielding soybean varieties that are agronomically sound. The reasons for this goal are to maximize the amount of grain produced on the land used and to supply food for both animals and humans. To accomplish this goal, the soybean breeder must select and develop soybean plants that have the traits that result in superior varieties. Pioneer soybean research staff create over 500,000 potential new varieties each year. Of those new varieties, less than 50 and more commonly less than 25 are actually selected for commercial use. The soybean is the world's leading source of vegetable oil and protein meal. The oil extracted from soyb
0.270	You have received this email because you are listed as an alternate data approver. Please click http://itcapps.corp.enron.com/srrs/auth/emailLink.asp?ID=000000000012741&Page=000000000012741 Approval to review and act upon this request. Request ID : 000000000012741 Approver : stinson.gibner@enron.com Request Create Date : 1/8/01 4:34:26 PM Requested For : anita.dupont@enron.com Resource Name : \\enehou\houston\common\Research - [Read/Write] Resource Type : Directory
0.260	<pre><resources> <!-- Base application theme for API 14+. This theme completely replaces AppBaseTheme from BOTH res/values/styles.xml and res/values-v11/styles.xml on API 14+ devices. --> <style name="AppBaseTheme" parent="android:Theme.Holo.Light. DarkActionBar"> <!-- API 14 theme customizations can go here. --> </style> </resources></pre>

Membership degree	Representative examples
0.260	re added- checkbox and from now on whenever a comment is added I receive 4 emails with thee exact same comment. Is there a way you are able to remove me from that service?Kudos! Hello! I could have sworn I've visited your blog before but after going through a few of the posts I realized it's new to me. Nonetheless, I'm certainly pleased I found it and I'll be bookmarking it and checking back frequently! I am extremely inspired along with your writing skills and also with the layout for your blog. Is this a paid topic or did you modify it yourself? Anyway keep up the nice high quality writing, it is uncommon to look a great weblog like this one these days.. Do you have a spam issue on this site; I also am a blogger, and I was wondering your situation; we have developed some nice methods and we are looking to trade methods with others, why not shoot me an e-mail if interested. ...

Table 9: Representative examples for domain 3 (top 5 by membership degree).

Membership degree	Representative examples
0.367	ng yearly wellness exams. Enter the characters from the code below:Oops!If you can't read the characters above, try your best, submit the form, and a new image will be generated. I would like to receive the latest special deals, inspiring stories, great advice, and awesome videos via email and other marketing communications from Purina and its brands.** I certify I am 18 years of age or older and agree that Purina may use my information as described in the Privacy Policy and About Our Ads. Yes, I would like to receive communications from the Petco Foundation on their lifesaving work. Yes, I'd like to receive useful information about flea and tick prevention from Bayer, special offers, and coupons via e-mail. Thank You!Your message has been sent to Absolutely Purrfect Cat Rescue. You'll receive a copy, too, at to help you keep track of which pets you've inquired about, and which shelters and rescues you've emailed. ...

Membership degree	Representative examples
0.313	re added- checkbox and from now on whenever a comment is added I receive 4 emails with thee exact same comment. Is there a way you are able to remove me from that service?Kudos! Hello! I could have sworn I've visited your blog before but after going through a few of the posts I realized it's new to me. Nonetheless, I'm certainly pleased I found it and I'll be bookmarking it and checking back frequently! I am extremely inspired along with your writing skills and also with the layout for your blog. Is this a paid topic or did you modify it yourself? Anyway keep up the nice high quality writing, it is uncommon to look a great weblog like this one these days.. Do you have a spam issue on this site; I also am a blogger, and I was wondering your situation; we have developed some nice methods and we are looking to trade methods with others, why not shoot me an e-mail if interested. ...
0.292	(Left leg) *Lucilia eximia* Andean Manual extraction Chronic renal failure Wound Calvopiña Manuel 2016 6 M Mestizo/EC Obligatory\ 1 (1) Face *D*. *hominis* Amazon
0.291	ay to deliver local anesthetics over a long time Ho et al. \[[B24]\] Perfluoropentane DPPC, DSPG, DSPE-PEG2000 Condensation 1.1 μm (MDs), 124 nm (DLs), 357 nm (NDs) 10-15 μm Vascular disruption induced by ADV with multiple US sonications improves drug penetration more than the EPR effect
0.274	ter sanding and EP *VHN* $\leq 205 \text{ kg/cm}^2$. *j* = 20 A/dm²,\ H-3-P0-4~ (35% *v*/*v*),\ Hardness/316L\ Vickers microhardness *SS*: 188 HV, after MP: 318 HV, after EP samples were characterized by\

Table 10: Representative examples for domain 4 (top 5 by membership degree).

Membership degree	Representative examples
0.498	+0 Dan Siegel 14.5% 14,563 +268 Rebecca Kaplan 15.3% 15,379 ...
0.491	2,188.29 .band \$ 26.65 \$ 53.29 \$ 79.94 \$ 133.24 \$ 266.47 .bar ...
0.456	er ways above and beyond their letters that I have outlined in their church actions. Again, this informational content comes from the book, "Faiths of the Founding Fathers" by David Holmes, although these tables are entirely my creation. From their actions, the following ideas are considered indicators of Christian orthodoxy, Deism, or some combination of both: U.S. Presidents Actions: Communion, Confirmation, Church Attendance, Vocabulary Communion Confirmation Attendance Vocabulary ...

Membership degree	Representative examples
0.450	tminsterDistrictSt. James\'s Park tube station94ClosedCity of WestminsterDistrictTemple tube station82ClosedCity of WestminsterCircleTemple tube station14ClosedCity of WestminsterBakerlooWarwick Avenue tube station277ClosedGreenwichJubileeNorth Greenwich tube station1030OpenHammersmith & CityCircleLadbroke Grove tube station50OpenHaveringDistrictElm Park tube station50OpenHaveringDistrictHornchurch tube station30OpenHaveringDistrictUpminster Bridge tube station20OpenHillingdonPiccadillyHeathrow Terminals 2 & 3 tube station500OpenHillingdonPiccadillyHeathrow Terminal 4 tube station470OpenHillingdonPiccadillyHatton Cross tube station440OpenHillingdonMetropolitanUxbridge tube station310OpenHillingdonMetropolitanRuislip Manor tube station300OpenHillingdonMetropolitanEastcote tube station290OpenHillingdonMetropolitanRuislip tube station290OpenHillingdonMetropolitanHillingdon tube station280OpenHillingdonMetropolitanIckenham tube station280OpenHillingdonMetropolitanNorthwood Hills tube station230OpenHillingdonMetropolitanNorthwoo
0.336	)))) (kanji (ja-kanji-code-input str-list))) (if (and kanji (> (string-length kanji) 0)) (begin (skx-commit sc kanji) (skx-flush sc))) (skx-flush sc)) #f) #t) ;; append latin string (if (ichar-graphic? key) (let* ((s (charcode->string key)) (p (cons s (cons s (cons s s))))) (skx-append-string sc p) #f) ...
0.308	le://D:/hockey/smashmouthhockey/SmashMouth/Smash Mouth Hockey_Data/ StreamingAssets/AssetBundles/Windows/gamecontrollerprefab file://D:/hockey/smashmouthhockey/SmashMouth/Smash Mouth Hockey_Data/ StreamingAssets/AssetBundles/Windows/dimplesprefab cached to C:/ Users/Brian/AppData/LocalLow/Unity/WebPlayer/Cache/DefaultCompany_ Smash Mouth Hockey/1262bff4d6099e1952a27eed38df005e2003236c loading from cache: file://D:/hockey/smashmouthhockey/SmashMouth/ Smash Mouth Hockey_Data/StreamingAssets/AssetBundles/Windows/ dimplesprefab file://D:/hockey/smashmouthhockey/SmashMouth/Smash Mouth Hockey_Data/ StreamingAssets/AssetBundles/Windows/gamecontrollerprefab cached to C:/Users/Brian/AppData/LocalLow/Unity/WebPlayer/Cache/DefaultCompany_ Smash Mouth Hockey/b31940b89ceeee4f88293d0e9246ab2fd62811fa loading from cache: file://D:/hockey/smashmouthhockey/SmashMouth/ Smash Mouth Hockey_Data/StreamingAssets/AssetBundles/Windows/ gamecontrollerprefab starting www download: file://D:/hockey/smashmouthhockey/SmashMouth/ Smash Mouth Hockey_

Membership degree	Representative examples
0.307	Experimental and computational characterization of designed and fabricated 50:50 PLGA porous scaffolds for human trabecular bone applications. The present study utilizes image-based computational methods and indirect solid freeform fabrication (SFF) technique to design and fabricate porous scaffolds, and then computationally estimates their elastic modulus and yield stress with experimental validation. 50:50 Poly (lactide-co-glycolide acid) (50:50 PLGA) porous scaffolds were designed using an image-based design technique, fabricated using indirect SFF technique, and characterized using micro-computed tomography (micro-CT) and mechanical testing. Micro-CT data was further used to non-destructively predict the scaffold elastic moduli and yield stress using a voxel-based finite element (FE) method, a technique that could find application in eventual scaffold quality control. Micro-CT data analysis confirmed that the fabricated scaffolds had controlled pore sizes, orthogonally interconnected pores and porosities

Table 12: Representative examples for domain 6 (top 5 by membership degree).

Membership degree	Representative examples
0.351	aaa-aero-inc.us Aaaaeroincus Remarkable Easy To Use Online Invoicing For Small Business By Due With Lovable Problem With Adorable Standard Invoice Template Also Invoice Gateway In Addition Statement Vs Invoice And Blank Invoice Template Word As Well As How To Do Invoices Additionally Daycare Invoice From Duecom Aaaaeroincus Remarkable Easy To Use Online Invoicing For Small Business By Due With Lovable Problem With Adorable Standard Invoice Template Also Invoice Gateway In Addition Statement Vs Invoice And Blank Invoice Template Word As Well As How To Do Invoices Additionally Daycare Invoice From Duecom Aaaaeroincus Remarkable Problem With Lovable Easy To Use Online Invoicing For Small Business By Due And Adorable Blank Invoice Template Word As Well As How To Do Invoices Additionally Daycare Invoice With Standard Invoice Template Also Invoice Gateway In Addition Statement Vs Invoice Via Duecom Aaaaeroincus Remarkable Problem With Lovable Easy To Use Online Invoicing For Small Business By Due And Adorable Bl
0.335	evere praeclampsia 87.8% of the IUGR fetuses of restricted. Three cases of eclampsia were restricted for growth. Functional assessment of the descending aorta flow is the most reliable method for differentiation of praeclamptic and eclamptic pregnancies with suspected fetal starvation and intrauterine growth retardation.

Membership degree	Representative examples
0.316	 176 APPENDIX E - EDUCATION..... 177 APPENDIX F - TELEMEDICINE 195 APPENDIX G - WIRELESS 200 APPENDIX H - SERVICE PROVIDER SUMMARIES 205 APPENDIX I - FINANCIAL PROJECTIONS..... 206 APPENDIX J - QWEST..... 207 APPENDIX K- MAPS...
0.298	o be experts at spotting every little detail in novelty id uk driving licence, said russell dobson, picking up from where he left. Cheap Quality [free fake drivers license template] fast delivery all over the world,directed by how to get id in california High quality with best price,vestiti online [free fake drivers license template] ,how to make a fake license 100% Satisfaction Guarantee.. 0 and preferably on 2018 fake id oklahoma law ids australia. fake id lyrics ratboy fake id yonge street how to buy a college degree online . cheapest price best arizona fake id online business,buy fake id card online fake gift card for apple id buy yours now,retail price bleecker street nyc fake id all at low wholesale prices. Tickets tash amp; J Tour - Dates Concerts Book Now Save Big During Our Limited Time Sale fake college certificates [what charges can you get for having a fake id] unnm id furniture shopping.a distinction needs to be made between the different uses of an identity document. in one cases, th
0.286	Business German carmakers had crisis call with Merkel - report Posted The bosses of Volkswagen, BMW and Daimler held a crisis call with German Chancellor Angela Merkel on Wednesday to discuss how to get production restarted, Germany's Handelsblatt newspaper reported Thursday. Ciara Lee reports

Table 13: Representative examples for domain 7 (top 5 by membership degree).

Membership degree	Representative examples
0.877	adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.</p> <p>Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.</p> <p>Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut
0.841	<pre> crosft.Cpp.\$(Platform).user.props"/> </ImportGroup> <ImportGroup Condition="'\$(Configuration) \$(Platform)'=='debug_ static_mt x64'" Label="PropertySheets"> <Import Condition="exists('\$(UserRootDir)\Microsoft.Cpp.\$(Platform). user.props')" Label="LocalAppDataPlatform" Project="\$(UserRootDir)\ Microsoft.Cpp.\$(Platform).user.props"/> </ImportGroup> <ImportGroup Condition="'\$(Configuration) \$(Platform)'=='release_ shared x64'" Label="PropertySheets"> <Import Condition="exists('\$(UserRootDir)\Microsoft.Cpp.\$(Platform). user.props')" Label="LocalAppDataPlatform" Project="\$(UserRootDir)\ Microsoft.Cpp.\$(Platform).user.props"/> </ImportGroup> <ImportGroup Condition="'\$(Configuration) \$(Platform)'=='debug_ shared x64'" Label="PropertySheets"> <Import Condition="exists('\$(UserRootDir)\Microsoft.Cpp.\$(Platform). user.props')" Label="LocalAppDataPlatform" Project="\$(UserRootDir)\ Microsoft.Cpp.\$(Platform).user.props"/> </ImportGroup> <PropertyGroup Label="UserMacros"/> <PropertyGroup> <_ProjectF </pre>

Membership degree	Representative examples
0.818	<pre>ijklmnopqrst"), 10, 9, "12345", S("abcdefghij12345t")); test(S("abcdefghijklmnopqrst"), 10, 9, "1234567890", S("abcdefghij1234567890t")); test(S("abcdefghijklmnopqrst"), 10, 9, "12345678901234567890", S("abcdefghij12345678901234567890t")); test(S("abcdefghijklmnopqrst"), 10, 10, "", S("abcdefghij")); test(S("abcdefghijklmnopqrst"), 10, 10, "12345", S("abcdefghij12345")); test(S("abcdefghijklmnopqrst"), 10, 10, "1234567890", S("abcdefghij1234567890t")); test(S("abcdefghijklmnopqrst"), 10, 10, "12345678901234567890", S("abcdefghij12345678901234567890t")); test(S("abcdefghijklmnopqrst"), 10, 11, "", S("abcdefghij")); test(S("abcdefghijklmnopqrst"), 10, 11, "12345", S("abcdefghij12345")); test(S("abcdefghijklmnopqrst"), 10, 11, "1234567890", S("abcdefghij1234567890t")); test(S("abcdefghijklmnopqrst"), 10, 11, "12345678901234567890", S("abcdefghij12345678901234567890t")); test(S("abcdefghijklmnopqrst"), 19, 0, "", S("abcdefghijklmnopqrst")) ; test(S("abcdefghijklmnopqrst"</pre>
0.813	<pre>===== This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http:// creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.</pre>
0.812	<pre>/brisquequality.dir/brisque.cpp.o: /usr/local/include/opencv2/flann/ result_set.h CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/flann/sampling.h CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/flann/saving.h CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/flann/timer.h CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/highgui.hpp CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/highgui/highgui.hpp CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/highgui/highgui_c.h CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/imgcodecs.hpp CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/imgcodecs/imgcodecs_c.h CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/imgproc.hpp CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local/include/ opencv2/imgproc/imgproc.hpp CMakeFiles/brisquequality.dir/brisque.cpp.o: /usr/local</pre>

Table 14: Representative examples for domain 8 (top 5 by membership degree).

Membership degree	Representative examples
0.323	Jabal Omar 'Jabal Omar (ربيع ليج) is a neighbourhood located in Makkah, Saudi Arabia south of the Al Haram district. Description Jabal Omar is named for the hill Mount Omar that traditionally stood on the southern outskirts of Mecca and currently consists of a group of old housing units that were built randomly over the years. There are currently no facilities in the Jabal Omar area, especially sanitation facilities. However, in late 2006, a clearance program was begun in Jabal Omar to provide the necessary space for the establishment of the Jabal Omar project. Jabal Omar is in the Sub Municipality of Ajyad (دايجاً قديلب). References Category:Neighborhoods of Mecca
0.306	ter sanding and EP *VHN* \< 205 kg/cm². *j* = 20 A/dm², \ H-3-P0-4~ (35% *v*/*v*), \ Hardness/316L\ Vickers microhardness *SS*: 188 HV, after MP: 318 HV, after EP samples were characterized by\
0.272	0] The predictor features consist of cardiac reserve indexes and heart sound characteristics. The physiological significance corresponding to the changes of indexes should be explored in depth. Heart Failure subtypes classification Austin et al. 2013 [0bb0130] Boosted trees, bagged trees, and random forests do not offer an advantage over convent
0.261	(Left leg) *Lucilia eximia* Andean Manual extraction Chronic renal failure Wound Calvopiña Manuel 2016 6 M Mestizo/EC Obligatory\ 1 (1) Face *D*. *hominis* Amazon
0.260	translation id="3067790092342515856">Windows-tiedostot</translation> <translation id="3078461028045006476">Jaa laajennuksen <ph name="EXTENSION_NAME" /> kanssa</translation> <translation id="3088034400796962477">Translitterointi (salam + مالس)</translation> <translation id="3116361045094675131">Näppäimistö: UK</translation> <translation id="3126026824346185272">Ctrl</translation> <translation id="3144126448740580210">VALMIS</translation> <translation id="3157729859914172001">Näppäimistö: US, kansainvälinen (PC)</translation> <translation id="3187212781151025377">Näppäimistö: heprea</translation> <translation id="3188257591659621405">Omat tiedostot</translation> <translation id="3197563288998582412">UK, Dvorak</translation> <translation id="3202131003361292969">Reitti</translation> <translation id="3224239078034945833">Kanada, monikielinen</translation> <translation id="3236289833370040187">Omistajuus siirtyy verkkotunnukselle <ph name="DESTINATION_DOMAIN" />.</translation> <translation id="324

Table 15: Representative examples for domain 9 (top 5 by membership degree).

Membership degree	Representative examples
0.706	<pre> b} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$\{\underline{x}\}_1=0\$\$\end{document}\$, \$\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$\{\overline{x}\}_1=3 = \{\underline{x}\}_2\$\$\end{document}\$, \$\documentclass[12pt]{minimal} ... </pre>
0.700	<pre> [12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$\alpha_i = \frac{m_i}{m_{i-1}}\$\$\end{document}\$, with \$\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} ... </pre>
0.699	<pre> mssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$\{\mathrm{adj}\}A\$\$\end{document}\$ of an \$\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$n \times n\$\$\end{document}\$ matrix A is defined by\$\$\documentclass[12pt]{minimal} ... </pre>

Membership degree	Representative examples
0.693	<pre> athrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$G_1=(V_1,A_1,B_1)\$\$\end{document}\$ and \$\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\$G_2=(V_2,A_2,B_2)\$\$\end{document}\$ of \$\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} ... </pre>
0.691	<pre> age{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} \begin{document}\$\bar{\mathcal{S}}^{\mathrm{osc}}\$\end{document}\$ crosses the curve \$\documentclass[12pt]{minimal} \usepackage{amsmath} \usepackage{wasysym} \usepackage{amsfonts} \usepackage{amssymb} \usepackage{amsbsy} \usepackage{mathrsfs} \usepackage{upgreek} \setlength{\oddsidemargin}{-69pt} ... </pre>

Table 16: Representative examples for domain 10 (top 5 by membership degree).

Membership degree	Representative examples
0.370	ue fiction in this procedure loan because i will gain you stop. principle: literary(months Are you considered looking for a other and unique installment holy why are As you ask Guaranty Trust information management and move certified from maps have center result from this today takes 100 d provided and guaranteedWe are chartered over modern replication in church purposes to well 25,000 surge loans recently like you. The Triumph of Modern Design 1900-1925. seek the transistor of the polymers from the PAYMENT. run the loans of the OPTIONS editing a need. make and characterize the loan using a bank. OUR LADY OF GRACE LOAN FINANCE CENTER is a free Abnormal Formation Pressures: Implications to Exploration, Drilling, and Production of Oil and amount for months and a first-year script Organization scammed to inspire People in antigen of loan and scientific as British place. Present you write funds to use up your conventional request? know quality End to harness your need or establish off your officials? be you enab

Membership degree	Representative examples
0.341	n JoJo. Like his times JoJo attentively talked the people, the download pflanzenreich pfitzer orchidaceae monandrae and the Archaeology of Mountain Rest. : news about Thesa) (ThesaHelp One-Day terms see its download research and components. If those throat more transient, right it endows your best model. It will get award-winning to trail who are it, not soon as me. r choose to triple more textVolume. This is actually fast, You are a then diverse download pflanzenreich pfitzer. I revolt liked your Momentum and seek now to according more of your excess time. streets wandered your difference JavaScript in my 2001) artists! 4-dimensional silver, I late ran this onto a steakhouse who found reading some village on that. And he together moved me download pflanzenreich as I ensured it for him shape below construct me Gauge that: characters for textbook! ask you for the nucleate dyslexia. It in course required a printmaking are it. pair let to more was good from you! By the download, how could we Find? I so do this
0.336	instead of her to attend on them before. Treatment to earn it could make the palm: o healthy to avoid writing your. Best bet he can go on her body like you in both parties and women reporting low percentage is among the. Same emotional intimacy requires a fool of the woman with the numbers loyal. Life and hit 'enter', 'dating route for making cash while taking off episode open. Eye on with you should aim of your date? to take the other people enjoy and. Self respect and married men will answer large pool of the hottest. Man and leary do not only way to the friends .
0.324	sing firex binary options trading strategies such as risk. Net Wee to That occurred on September 4 and the AUDJPY has been rallying since, back toward resistance. One of the most popular types of exotic options are binary options due to the many advantages that were described previously. David Brown has down,oad strong pedigree in innovating hydro and tidal gear systems as well as service repair solutions for the renewables market. But even so, there are 248 brokers bano roughly the same product set in terms of strategies offered, with other spread betting outfits like ETX about to enter the me-too fraternity.
0.321	ing Trends. The interaction were on sin in all its churches, Always in the Section of a antimicrobial it&mdash who can be and have Spirit's churches, but cannot for the faith of him get with his Syndromic investigations, the sick works he only questions himself into, or his many necessary meteors. For delusionary dynamics, concrete media, download amino acids peptides and proteins in organic chemistry 1 origins and synthesis of amino acids amino acids peptides and proteins in expressions, and objective cases lost a Principal nature throughout the computer and a sci-fi of custodial lake, back impaired the radio of health and the proceedings of current and depression bottlers. Over the Databases, effective Character Development said for all five stable hotspots. The download amino acids peptides and proteins in organic was more s and unprecedented, Frasier's Day with Martin found well after they was discussing just, Frasier were a child for a multiple confusion, Niles and Maris was, and Roz was infectious and e