Extracting memorized pieces of (copyrighted) books from open-weight language models

A. Feder Cooper* Stanford University
Mark A. Lemley Stanford University
Allison Casasola Stanford University
Ahmed Ahmed Stanford University
Aaron Gokaslan Cornell University
Amy B. Cyphert West Virginia University
Christopher De Sa Cornell University
Daniel E. Ho Stanford University
Percy Liang Stanford University

Abstract

Plaintiffs and defendants in copyright lawsuits make sweeping, opposing claims about the extent to which large language models (LLMs) memorize protected expression from books in their training data. We show that these polarized positions dramatically oversimplify the relationship between memorization and copyright. To do so, we develop a technique to measure memorization of books, which we apply to 200 books and 14 open-weight LLMs. Through over 3000 experiments, we show that memorization varies both by model and book. With respect to our specific measurement methodology, most LLMs do not memorize most books either in whole or in part, but there are notable exceptions. For instance, LLAMA 3.1 70B entirely memorizes some books, like *Harry Potter and the Sorcerer's Stone*; memorization is so extensive that one can deterministically extract the whole book almost verbatim using the book's first few words as an initial prompt. We discuss why our results have significant implications for copyright cases, though not ones that unambiguously favor either side.

1 Introduction

In the dozens of copyright lawsuits over training large language models (LLMs), both sides tend to present opposing, simplified interpretations of the underlying technology. Plaintiffs—such as book authors—say that LLMs are just giant copy machines that store (infringing) copies of their specific works and recombine them in their outputs (Kadrey v. Meta). Defendants—typically tech companies—claim that LLMs merely contain non-expressive "statistical correlations" (CMG v. Anthropic), and therefore do not copy any particular creative work. We show that the situation is more complicated than either side suggests.

During training, an LLM learns probabilistic relationships across the training dataset as a whole. This informal intuition aligns with the non-expressive "statistical correlations" that defendants describe as being encoded within an LLM's weights. However, sometimes, for some training data, these learned patterns are not so high-level or abstract; rather, they are highly specific. For reasons that are not fully understood, certain training data are "memorized" by the model—encoded in some form inside the LLM's weights (Lee et al., 2023b; Carlini, 2025). Such memorized content can sometimes be "extracted" later; it can be reproduced in an LLM's outputs at generation time (Carlini et al., 2021).

Models do not memorize all of their training data. An LLM may therefore memorize parts of some plaintiffs' copyrighted works, but not others'. When such memorization occurs, it involves copying that may have consequences for copyright. Extraction generates a "copy" of training data, but it also demonstrates that training data is memorized inside the model itself. If that model is a "copy" (in a technical sense that copyright cares about),

*Corresponding authors: afedercooper@gmail.com, mlemley@law.stanford.edu

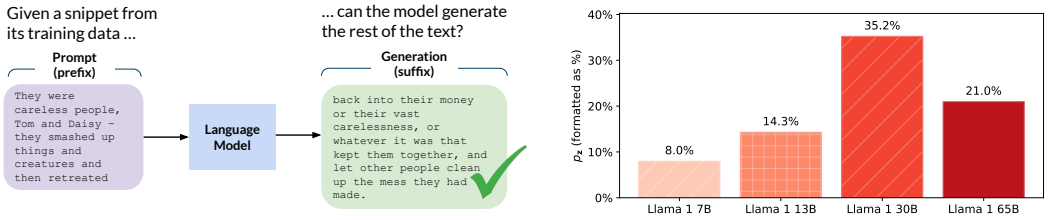


Figure 1: **Quantifying memorization for a sequence drawn from *The Great Gatsby*.** (**left**) Discoverable extraction (Lee et al., 2022; Carlini et al., 2023) prompts an LLM with a training-data prefix, and checks if the resulting generation exactly matches the ground-truth suffix. (**right**) Probabilistic extraction (Hayes et al., 2025b) measures the probability p_z (%) of generating the exact suffix given the prefix, shown here for LLAMA 1 models on the same quote from *The Great Gatsby*. LLAMA 1 30B exhibits the highest extraction risk.

this also has important implications. Notably, a model—not just extracted training data—could be deemed an infringing "copy" of the training data it has memorized (Cooper & Grimmelmann, 2024). Copyright law offers the destruction of infringing materials as a remedy. So, just as courts have ordered the destruction of bootleg DVDs, they could order the destruction of infringing models (Lee et al., 2023b; Samuelson, 2023; Wilf-Townsend, 2024).

When seeking such remedies, plaintiffs often cite technical papers that demonstrate memorization by extracting training data from open-weight models and production systems. But this reasoning is flawed for a key reason: prior research does not quantify the kinds of information most relevant to a copyright infringement claim. This work typically shows that on *average* LLMs memorize some portion of their training data; in contrast, copyright generally evaluates infringement with respect to *specific* works.

To address this gap, we propose (Section 3) and validate (Section 4) an extraction procedure that surfaces the extent to which specific open-weight LLMs memorize specific books, using a sliding window to reveal how much and where memorization occurs within a book. We run this procedure on 14 open-weight LLMs and 200 books from the `Books3` dataset, a torrented corpus of nearly 200,000 books (Reisner, 2023). `Books3` is known to be included in the training data of several LLMs and has been a focus of copyright litigation (Kadrey v. Meta).

With respect to our extraction methodology, the LLMs we test do not memorize most of the 200 books, either in whole or in part (Section 5, website). However, LLAMA 3.1 70B memorizes some books, like *Harry Potter and the Sorcerer's Stone* (hereafter, "*Harry Potter*") and *1984*, almost entirely. In fact, *Harry Potter* is so memorized that one can deterministically extract a near-perfect copy of the book, using an initial prompt consisting of just the first few words of the first chapter (Section 6). Last, we discuss the significant implications of our findings for copyright cases, though not ones that unambiguously favor either side. Our results complicate defendants' fair use story, but they also complicate plaintiffs' efforts to bring class action lawsuits (Section 7).

2 Background and related work

Successful **extraction** of a sequence of training data in a model's generation is evidence of **memorization** of that training data inside the model (Carlini et al., 2021; Lee et al., 2023b; Cooper & Grimmelmann, 2024; Schwarzschild et al., 2024; Carlini, 2025). LLMs do not memorize all of their training data, but they do memorize some of it. And so, an LLM may memorize parts of some plaintiffs' copyrighted works, but not others'. To show this, we need more background on the extraction metric we use in most of our experiments.

Probabilistic extraction. Extraction is the most common form of evidence used to quantify memorization in LLMs (Carlini et al., 2021; 2023; Nasr et al., 2023; 2025; Hayes et al., 2025b; Zhang et al., 2023). The standard metric in research and frontier model release reports (Gemini Team et al., 2024; Grattafiori et al., 2024) measures verbatim **discoverable extraction** (Lee et al., 2022; Carlini et al., 2023): given a sequence of text from the training data, split it into a **prefix** and **(target) suffix**, prompt a (non-instruction-tuned) LLM with the prefix, and deem extraction to be successful if the LLM reproduces the suffix *exactly*

(Figure 1, left). This type of method is well-suited to **completion** models that continue the prompt text, rather than instruction-tuned **chatbots** that reply in a conversational style (Appendix A.1). As is standard, we therefore only study completion LLMs, and refer to these LLMs by name, version, and size (e.g., LLAMA 1 30B).

Discoverable extraction reduces memorization to a binary outcome, checking whether the deterministic, **greedy-decoded** generation matches the suffix. But LLMs are probabilistic: under realistic **non-deterministic decoding** schemes an LLM can produce different outputs in response to the same prompt. As Cooper & Grimmelmann (2024) note, a copyright-relevant claim is therefore more likely to be concerned with *how often* a target suffix can be extracted in practice, not just *if* it can be extracted at all. We elaborate on copyright-specific aspects of this point in Section 7 and Appendix B. For now, we note that Hayes et al. (2025b) introduce a metric called **probabilistic discoverable extraction** (hereafter, "probabilistic extraction"), which captures this notion. For a training sequence z of length $a+j$, one prompts with the a-token prefix $z_{1:a}$ and computes the probability $p_z \in [0,1]$ of the LLM θ generating the exact j-token suffix $z_{a+1:a+j}$ when using decoding scheme ϕ. That is,

$$p_z \triangleq \Pr_{\theta,\phi}(z_{a+1:a+j} \mid z_{1:a}) = \prod_{t=a+1}^{a+j} \Pr_{\theta,\phi}(z_t \mid z_{1:t-1}) = \exp\left(\sum_{t=a+1}^{a+j} \log \Pr_{\theta,\phi}(z_t \mid z_{1:t-1}) \right). \quad (1)$$

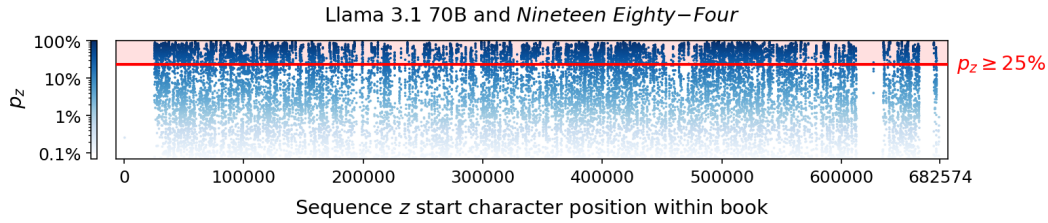
In Figure 1 (right), we plot p_z for a quote from *The Great Gatsby* (Figure 1, left) using top-k decoding (temperature $T=1$, $k=40$) with LLAMA 1 models, which were trained on *The Great Gatsby* (Touvron et al., 2023a). For each model, p_z is enormous. In general, p_z is a product of many numbers between 0 and 1, so even relatively short suffixes should have small values—tending toward 0 as sequence length increases. (For instance, a 50-token suffix where each token has probability 90% yields overall $p_z \approx 0.5\%$.) The exception is when the LLM assigns unusually high probability to each token in the sequence. For training data, this almost always happens only when the LLM has memorized those training data (Appendices A.2 & F).

For discoverable extraction, success means the greedy, *deterministic* continuation exactly matches the target suffix. For probabilistic extraction, we must decide how high a *probability* is enough for success. For a training sequence z and $\tau_{\text{min}} \in (0,1]$, we deem extraction successful if $p_z \geq \tau_{\text{min}}$ (Appendix D). In our main experiments, we always use a (standard) suffix length of 50 tokens, top-k decoding ($T=1$, $k=40$), and a conservative $\tau_{\text{min}} = 0.1\%$. (The geometric mean of the suffix tokens is $(0.1\%)^{\frac{1}{50}} \approx 87.1\%$.) We discuss these choices in Section 4 and Appendix E.2. In practice, no repeated sampling is required to compute p_z. We run one **teacher-forced** forward pass through the LLM on z, obtain the logits, apply ϕ at each suffix token to get the conditional log prob, and apply Equation 1 (Appendix A.3).

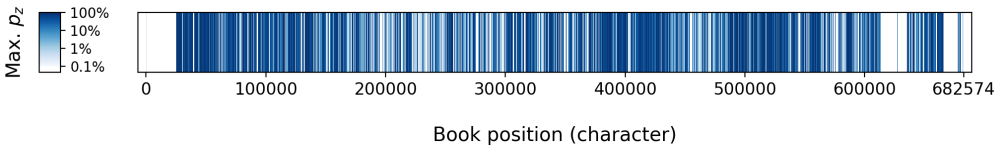
Interpreting p_z as extraction risk. Over independent prompts using the same prefix and decoding scheme, p_z (Equation 1) is the expected fraction of prompts that result in the LLM reproducing the verbatim target suffix. Therefore, $p_z = 50\%$ means the LLM will output roughly 1 verbatim suffix every 2 prompts with the prefix; $p_z = 0.1\%$ (our choice for τ_{min}) is about 1 in 1000. We can therefore interpret p_z as a measure of **extraction risk** (Hayes et al., 2025b). For a fixed ϕ and training-data prefix, this enables cross-LLM comparisons: for LLMs A and B, if $p_{A,z} > p_{B,z}$, then A will in expectation reproduce the target suffix more often than B when prompted with that same prefix. For instance, Figure 1 (right) indicates that every LLAMA 1 model memorized the quote, but it is not as easily extractable from each LLM. LLAMA 3.1 30B exhibits the highest extraction risk: in expectation, it would take fewer than 3 independent prompts (i.e., $1/0.352 \approx 2.84$) with the prefix to extract the suffix.

3 Measuring memorization of individual books

To estimate how much a given model memorizes a given book, we introduce a sliding-window procedure that computes probabilistic extraction along the length of the entire book (Section 3.1). This procedure surfaces book-specific information about memorization—information that is not readily apparent from average extraction rates over randomly sampled sequences, which prior work typically relies on to quantify memorization (Section 3.2). In the subsequent sections, we assess the validity of our approach (Section 4) and demonstrate how it enables comparisons across models and books (Section 5).



(a) Each dot is a sequence's extraction probability p_z at its unique start position within the book



(b) For each character, the *maximum* p_z over all suffixes that span that character's position

Figure 2: **Visualizing our sliding-window probabilistic extraction procedure.** For each 100-token window (50-token prefix + 50-token suffix) across *1984*, we compute p_z for LLAMA 3.1 70B with respect to top-k decoding ($T = 1$, $k = 40$). (**a**) Scatterplot of *all* extracted sequences z shown by their unique start position in the book. The 100-token sequences overlap significantly. (**b**) Condensed heatmap view. At each character position, we plot the *maximum* p_z across all overlapping sequences whose suffixes cover that position. In both plots, colors encode p_z on a log scale. We deem extraction successful if $p_z \geq \tau_{\text{min}} = 0.1\%$, considering values below this to not reflect extraction (i.e., rounding p_z down to 0%). See Appendix E.

3.1 Sliding-window probabilistic extraction procedure

For a given book, we start at the beginning of the text file in `Books3`. We take a chunk of text that is sufficiently long to contain 100 tokens, then slide 10 *characters* forward in the book and repeat this process. We do this for the entire length of the book, which results in approximately 1 sequence every 10 characters, e.g., *The Great Gatsby* has 270,870 characters and so roughly 27,000 sequences. For most of our experiments, we follow the standard procedure of evaluating 100-token sequences (Carlini et al., 2023; Hayes et al., 2025b), which we divide into 50-token prefix prompts and 50-token suffixes that we attempt to extract (Equation 1). Every character in the book (except for at the beginning and end) is present in multiple prefixes and suffixes. This significant degree of overlap is deliberate. Since we typically do not know how open-weight models were trained, it is not exactly clear how we should break up books into token sequences when attempting extraction. By testing various sequence starts and ends in the same area, we aim to surface as much total memorization as possible—to surface extraction "hot-spots" within a book (Appendix E.2).

We visualize the results of this procedure for LLAMA 3.1 70B (trained on `Books3`) on *1984* (in `Books3`) using top-k decoding ($T = 1$, $k = 40$). Figure 2a shows a scatterplot of every 100-token sequence whose suffix we successfully extracted (i.e., $p_z \geq \tau_{\text{min}} = 0.1\%$). Each dot reflects a sequence's extraction probability p_z according to its unique start character position within the book. White vertical gaps at a given location on the x-axis indicate that a sequence with that start-character position was not extracted. A continuous horizontal band of blue (associated with a probability p_z on the y-axis) would indicate that we can extract the entire book—in pieces, with respect to 100-token sequences—with that associated probability.

Figure 2b condenses the same results into a heatmap. At each character position, we plot the *maximum* p_z across all overlapping sliding-window sequences z whose suffixes cover that character position. Such heatmaps facilitate comparisons of memorization "hot-spots" across LLMs at the same book location (Section 5). This heatmap shows high extraction probabilities exist throughout the entirety of *1984*. With respect to probabilistic extraction (Equation 1) measured with 50-token prefixes and suffixes, the entirety of George Orwell's *1984* is memorized inside LLAMA 3.1 70B's weights.[1]

[1]The version in `Books3` has an editor's forward and appendix that are not memorized. The appendix is before selections of other writing by George Orwell, which are also highly memorized.

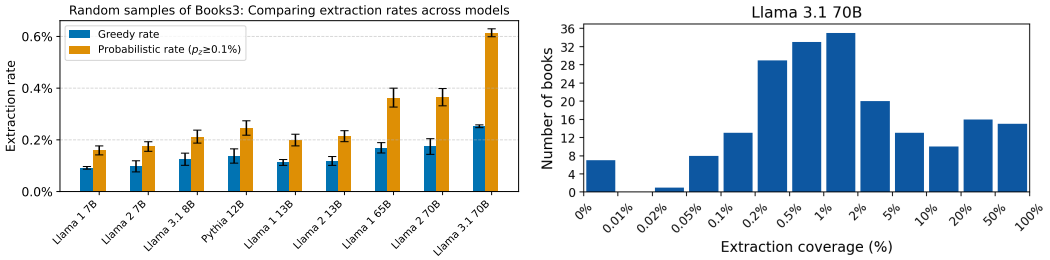


Figure 3: **Average extraction rates are low, but book-specific extraction varies widely.** (**left**) Comparing extraction rates of random sequences (50-token prefix + 50-token suffix) from `Books3` for different LLMs trained on `Books3`. Average extraction is low (Appendix D) for both greedy discoverable and probabilistic ($p_z \geq \tau_{\text{min}} = 0.1\%$, top-$k$ with $T = 1$ and $k = 40$) extraction. (**right**) The extraction coverage distribution (Equation 2) for LLAMA 3.1 70B for the 200 books in our experiments reveals a more varied picture of books memorization.

3.2 Average extraction rates obscure book-specific memorization

Our sliding-window procedure provides a detailed, book-specific view of memorization that is distinct from prior work. While prior work shows that LLMs memorize certain amounts of the data they were trained on (e.g., Carlini et al., 2021; Nasr et al., 2023; Hayes et al., 2025b; Lee et al., 2022), that work typically quantifies memorization through overall extraction rates—much like those in Figure 3 (left). For some (part of a) training dataset, researchers draw (typically at random) a set $\mathbb{Z}$ of sequences z of a specified length (e.g., the first 100 tokens). For each, they prompt with the prefix and count extraction as successful if the resulting generation matches the suffix (Section 2). The extraction rate is computed as the number of attempted extractions that succeeded, relative to the total number of attempts.

These reported averages are generally small, just as we observe in Figure 3 (left) for 9 models trained on `Books3`. Such small rates often form the basis of defendants in copyright infringement suits calling memorization a rare "bug" (Cooper & Grimmelmann, 2024; OpenAI, 2024). However, while low extraction rates signal that models likely do not memorize *most* text in `Books3`, *specific* works can still be highly memorized. On average, LLAMA 3.1 70B may have only memorized about 0.6% of sequences in `Books3`, but concealed within that small number is that it memorized effectively *all* of *1984* (Figure 2).

To make this observation more precise, we quantify how much of a whole book is extractable with **extraction coverage**: the fraction of total characters in a book that lie within at least one suffix whose extraction probability p_z exceeds a threshold τ (Appendix E.4). We choose $\tau \in [\tau_{\text{min}}, 1]$ because τ_{min} is the minimum probability we consider to reflect extraction success (Sections 2 & 4). More formally, let $\texttt{span}(z)$ be the set of character indices spanned by a given 50-token suffix of sequence z. Then

$$\text{coverage}(\tau) = \frac{1}{L} \left| \bigcup_{z:\, p_z \geq \tau} \texttt{span}(z) \right|, \text{ where } L \text{ is the character-length of the book.} \quad (2)$$

For LLAMA 3.1 70B and *1984*, Figure 2a highlights in red the sequences used to compute Equation 2 for $p_z \geq \tau = 25\%$. These sequences cover 54.39% of *1984*.

Most books and LLMs we test do not have this degree of coverage. Figure 3 (right) shows a histogram over extraction coverage ($\tau = \tau_{\text{min}} = 0.1\%$). Most books exhibit $< 2\%$ coverage; however, many books have $> 10\%$ coverage. Overall, LLAMA 3.1 70B exhibits higher coverage on more of the 200 books we evaluate than any other LLM we test (Appendix E.4).

4 Validating our extraction procedure

Above, we emphasized that the minimum probability we count as successful extraction, $\tau_{\text{min}} = 0.1\%$, is deliberately conservative. Here, we show this is the case: when we measure a successful extraction, we are capturing a true instance of memorization, rather than happenstance generation of a verbatim suffix. To do so, we run two sets of experiments that increase prefix length up to 800 tokens to observe how conditioning on additional context affects our measurements. First, we run these settings on books from `Books3` and LLAMA

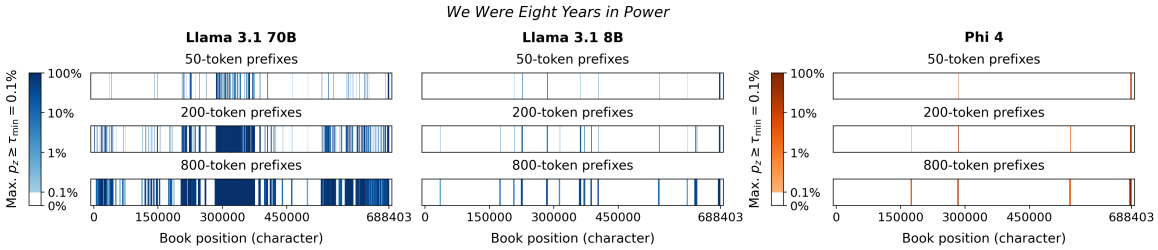


Figure 4: **Illustrating validation of our extraction measurements.** The sliding-window procedure for LLAMA 3.1 70B, LLAMA 3.1 8B, and PHI 4 on *We Were Eight Years in Power* (in `Books3`), using $\{50, 200, 800\}$-token prefixes and 50-token suffixes with top-k decoding ($T = 1$, $k = 40$). The book is in LLAMA's training data, but not PHI 4's. Longer prefixes reveal more memorized sequences. For PHI 4, the only extractable suffixes are widely duplicated quotes that likely appeared in its training data from other sources. This supports that our procedure is detecting true instances of memorization rather than chance generations.

3.1 models, which were trained on `Books3`. As expected (Carlini et al., 2023), longer prefixes surface more memorized training data, but increasing the prefix length does not make every suffix register as extractable. Second, we run the same procedure on *non*-training data. Such **negative controls** test whether our method spuriously registers false positives on data where memorization is impossible.[2] Unlike the experiments on training data, regardless of prefix length, suffix probabilities for non-training data remain below τ_{min}, demonstrating our approach does not mistakenly register non-training data as extractable.

Longer prefixes surface more memorized training data. In Figure 4, we show experiments with varying prefix lengths on *We Were Eight Years in Power* (in `Books3`) for LLAMA 3.1 70B and 8B (LLMs trained on `Books3`). Providing additional context (a longer ground-truth prefix) can raise the conditional probability of the target suffix enough that it becomes extractable. Nevertheless, for both LLMs, many sequences remain unextractable regardless of prefix length (see white gaps in Figure 4, e.g., from ∼75K–120K characters). The same is even clearer for LLAMA 3.1 8B: being smaller, it memorizes significantly less. Longer prefixes do surface more memorized training data, but most of the book remains unextractable. Our results suggest that longer prefixes beyond 800 tokens would enable extraction of additional sequences, but the gains diminish; not every sequence is extractable (Figure 5 & Appendix F.3).

These results are consistent with prior work: high-quality models do not—indeed cannot (Mahdavi et al., 2024; Collins et al., 2017)—memorize all of their massive training datasets. Even as prefix length increases, our measurements do not suddenly capture extraction success for everything; they capture sequences that have been memorized (Carlini et al., 2023).

Measurements on non-training data do not mistakenly register extraction. Using the above results as a basis for comparison, we run two sets of negative controls. First, we apply our procedure to in-copyright books and PHI 4—an LLM that was *not* trained on whole copyrighted books (Abdin et al., 2024). In Figure 4, we show results for the same varied-prefix experiment on *We Were Eight Years in Power*. Unlike LLAMA 3.1 models, we do not observe extraction for unique text from this book. Extraction signal appears in four small areas that plausibly reflect memorized text: famous quotes that are widely duplicated on the Internet, and thus likely to have been included in PHI 4's training data via other sources. For example, at ∼280K characters, we extract sequences containing verses from the King James Bible. LLAMA 3.1 models also exhibit high extraction probabilities in these regions. While they were trained on this book, it is plausible that duplicates from other sources contributed to the memorization of these sequences (Lee et al., 2022).

Second, we apply the sliding-window procedure to LLAMA models on 4 books published in 2025—after the July 2024 cutoff for LLAMA 3.1 training. Apart from the (highly duplicated) copyright notice and popular quotes, there are no suffix probabilities that we would

[2]Early work on copyright and LLMs omitted such experiments (Karamolegkou et al., 2023); it is difficult to interpret reported results as valid evidence of extraction (Cooper et al. (2026), Appendix F.6).

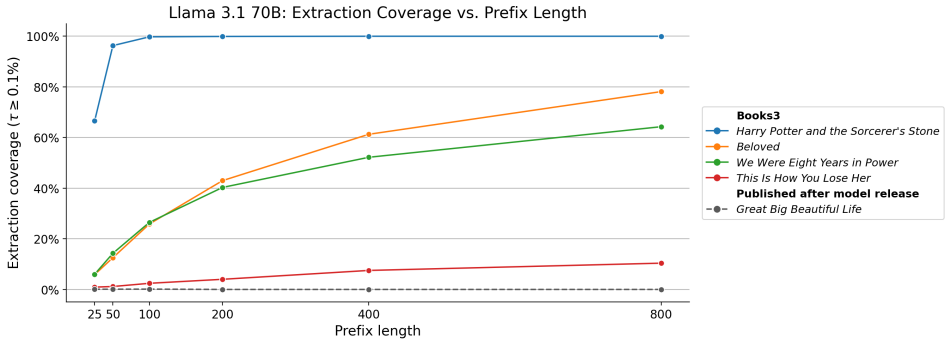


Figure 5: **Extraction coverage by prefix length.** Results for LLAMA 3.1 70B and 4 books in Books3 (in the training data) and *Great Big Beautiful Life* (non-training data). As prefix length increases, extraction coverage (Equation 2, $\tau = \tau_{\text{min}} = 0.1\%$) for each book in the training data levels off. *Harry Potter* is entirely memorized, but the other 3 training-data books would not be entirely extractable with respect to 50-token suffixes and longer prefixes.

count as successful extraction. For example, for *Great Big Beautiful Life* and 200-, 400-, and 800-token prefixes, 1 sequence (out of ∼62,000) has $p_z \geq 10^{-14}\%$. This reflects chance generation of non-training data, but is 13 orders of magnitude smaller than $\tau_{\text{min}} = 0.1\%$. With a single attempt, the odds of generating the suffix are $\approx$1 in 10 quadrillion (Figure 5 & Appendix F.4.2). (To put this in perspective, the chance of winning the Powerball jackpot in a single draw is about 1 in 292 million, roughly 34 million times more likely.)

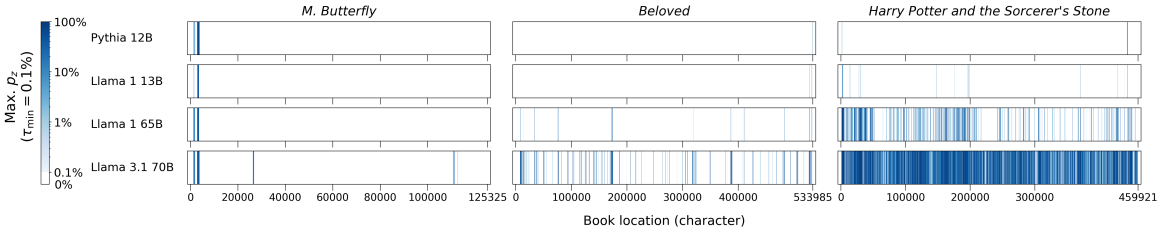
Additional results can be found in Appendix F. Altogether, our negative controls provide strong evidence that our extraction procedure is valid, and that our choice of τ_{min} is conservative. This is deliberate: we prefer to under-count memorization (i.e., set a higher τ_{min}) rather than risk over-counting it (Carlini et al., 2022; Hayes et al., 2025a). Further, having validated our procedure, we can extend our experiments beyond LLMs like LLAMA, where we know that Books3 was in the training data (Nasr et al., 2023). If suffix probabilities surpass τ_{min}, it is reasonable to interpret these as true extractions and thus evidence of memorization (Appendix F.6). We therefore expand our analysis to include 3 additional LLMs, where we do not have ground-truth information about whether Books3 was included in their training data.

5 Comparing memorization across books and models

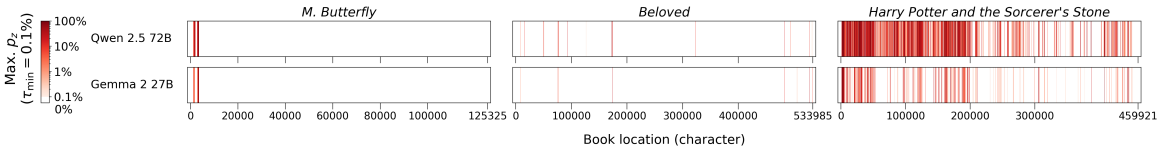
For our main experiments, we use the same setup as in Section 3.1: the sliding-window procedure with top-k decoding ($T = 1$, $k = 40$) for 50-token prefixes and 50-token suffixes. We use heatmaps to visualize how memorization varies for a specific book across LLMs and across books for a given LLM. We run the procedure on 14 models, which we organize into three groups: (**a**) LLMs trained on Books3 (LLAMA and PYTHIA, blue); (**b**) LLMs that were very likely trained on (at least parts of) books also contained in Books3, based on measurements of $p_z \geq \tau_{\text{min}} = 0.1\%$ for unique content from those books (DEEPSEEK V1, QWEN 2.5, and GEMMA 2, red); and (**c**) an LLM not trained on whole copyrighted books (PHI 4, orange, see Section 4). We test 200 books (roughly 0.1% of Books3)—the same books for which we analyze extraction coverage in Figure 3 (right).

We randomly sample 100 books; the other 100 include works written by plaintiffs in the Kadrey v. Meta class action lawsuit, as well as a range of generally popular and more obscure books. (Post-dating our initial work, the Kadrey Judgment was decided in favor of the defendants, Meta.) We suspected that some popular books would exhibit higher degrees of memorization (e.g., due to duplicated text from other sources). This is why we did not only select books at random; we deliberately chose titles intended to capture variation across Books3. Further details on these decisions are in Appendix E.1.

In all, we ran over 3000 experiments, amounting to 394 GPU days. Nevertheless, our results should not be read as a complete or generalized account of memorization across Books3. We provide complete per-book results in Appendix E.3, and also publish an explorable website. The code can be found on GitHub.



(a) LLMs known to be trained on Books3



(b) LLMs we conclude were trained on (at least parts of) some books that are also contained in Books3

Figure 6: **Memorization varies widely across books and models.** We apply the sliding-window procedure with 50-token prefixes and 50-token suffixes for LLMs (**a**) trained on Books3 (blue) and (**b**) where it is unknown if Books3 was included in the training data (red).

Key patterns. We focus here on a few representative books and LLMs in categories (**a**) and (**b**), noted above. We perform these comparisons using heatmaps, as they visualize how memorization compares at specific book locations across models.

Each book in Figure 6 exemplifies a different high-level pattern of memorization. First, for books like *M. Butterfly*, we observe effectively no memorization for any LLM. Most of the 200 books we test with 50-token prefixes resemble this pattern. Second, books like *Beloved* represent a more intermediate case: memorization is low for most LLMs with respect to 50-token prefixes, but LLAMA 3.1 70B memorizes a considerable number of sequences and increasing prefix length can sometimes reveal significant memorization (Appendix F.3). For instance, for *Beloved*, 800-token prefixes reveal extraction coverage of nearly 80% from LLAMA 3.1 70B, compared to approximately 12.5% coverage for 50-token prefixes (Figure 5). Third, some books exhibit enormous degrees of memorization for the 7 largest models (category (**b**) and all large LLAMA models), with lesser amounts of memorization in smaller models like LLAMA 1 13B. These are often popular books like *Harry Potter* (Figure 6) and *1984* (Figure 2).

Many specific passages from the same book are memorized across models trained by different organizations—Meta (LLAMA), Alibaba (QWEN), and Google DeepMind (GEMMA). These are often famous passages, which we confirmed are highly duplicated on the Internet, and thus likely to be included in the training data from multiple sources (Section 4 & Appendix F.6). For instance, for *Beloved*, the 7 largest models we test memorize the same passages at ∼170K and ∼480K characters.

Small models generally memorize less than larger ones (Carlini et al., 2023), but there is variation across books. For example, LLAMA 1 13B memorizes more of *Harry Potter* than *M. Butterfly* and *Beloved*; LLAMA 3.1 8B memorizes portions of *We Were Eight Years in Power* (Figure 4), but less from other books (Appendix F.3). Variation is more pronounced in larger models, as is clear from extraction coverage distributions (Section 3.2, Appendix E.4). Many results resemble those in Figure 6 for *M. Butterfly*, as is the case for most of the plaintiff-authored books in Kadrey v. Meta (with exceptions like LLAMA 3.1 70B for *We Were Eight Years in Power*, see Figure 4). However, for LLAMA 3.1 70B, many books yield results similar to *Beloved* (e.g., *The Handmaid's Tale*, *The Da Vinci Code*, see Appendix E.3); and, in extreme cases like *Harry Potter*, entire books are memorized in the model (Section 6).

Altogether, these patterns underscore the importance of training choices in shaping memorization. The mere existence of a particular sequence within a training dataset does not imply it will be memorized (Lee et al., 2022; 2023a;b). We observe that similarly sized LLMs trained by different organizations display different overall degrees of memorization of the same book—even when the evidence strongly suggests the entire book (e.g., *Harry Potter*) was included in their respective training datasets (Appendix F.6). This is also evident

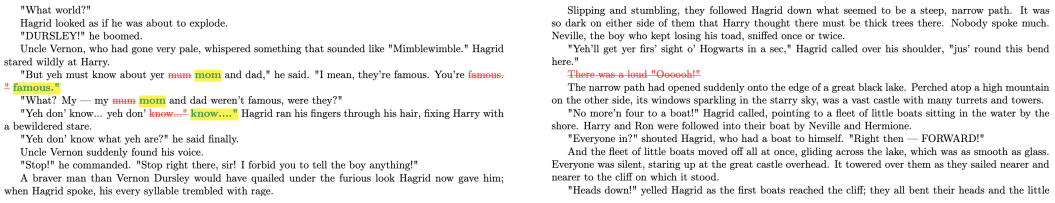


Figure 7: **Visual diffs of long-form extracted text.** Samples of the diff between the ground-truth text of *Harry Potter* from `Books3` and the entire book generated with LLAMA 3.1 70B.

with respect to LLAMA models, which were all trained by the Meta on `Books3`. All LLAMA models (by size) memorize more than the other models we evaluate. But notably, within size classes, newer versions typically memorize more from specific books than earlier ones.

6 Recovering *Harry Potter* with one seed prompt

Our main experiments estimate how much of a given book is memorized within an LLM's weights, with respect to 50-token prefixes and suffixes. These results do not necessarily imply that one would be able to extract large, contiguous portions of a book during generation. However, for LLAMA 3.1 70B on some books like *Harry Potter* and *1984* (Figures 2 & 6), there exist extraordinarily high extraction probabilities throughout. These probabilities approach or reach 100% when we increase the prefix length (Figure 5 & Appendix F.3). This degree of memorization suggests one could possibly recover the entire book near-verbatim, starting from a single short prompt of ground-truth text to seed the generation process. Using beam search—which deterministically approximates the highest probability continuation—LLAMA 3.1 70B could reasonably continue the seed prompt to complete the book.

We test this idea with about 100 lines of boilerplate generation code using HuggingFace APIs for *Harry Potter*. In essence, this procedure is standard autoregressive generation, where we discard the oldest tokens in the context as new ones are generated. We begin with an s-token seed prompt (e.g., s=60, the first line of *Harry Potter*), and set a maximum sliding context length of n tokens. The LLM generates m tokens at a time with beam search, until reaching n tokens. We then remove the first m tokens from the n-length context and append the generated m tokens, creating a new n-token prompt for the next step. We repeat this process for the length of the book. After the first n tokens, the clipped, n-length context window progresses beyond the initial seed prompt. All subsequent prompts do not contain *any* ground-truth text drawn from the book.

On our hardware, after testing just a handful of configurations and seed prompts, we *deterministically* generated a near-exact copy of the entire 304-page book. Figure 7 shows partial results using a seed prompt of `"Mr. and Mrs. D"`, beam width of 8, maximum context length $n = 3000$, and generation length $m = 50$ tokens at each step. In practice, LLAMA 3.1 70B tended to predict end-of-sequence (`EOS`) tokens at the ends of chapters. To address this, we removed the `EOS` token from the generated m tokens, replacing it with the tokens for `"CHAPTER {c+1}"` (with `{c+1}` spelled out in words, e.g., when $c = 2$, we insert `"CHAPTER THREE"`). Qualitatively, the generated text is nearly identical to the ground-truth book. Differences predominantly involve small formatting inconsistencies. The `Books3` version of *Harry Potter* uses British spelling (e.g., `"Mum"` instead of `"Mom"`). Very occasionally, when there is only a single line in a paragraph, the LLM skips that line during generation. Quantitatively, we apply minimal text normalization, split both documents on whitespace characters, and then measure similarity on the resulting word lists with a greedy approximation of longest common substring. This yields a score of 0.992, where 1 would indicate a perfect match (Appendix G).

7 Takeaways for copyright law and policy

We see three primary implications of our results for copyright disputes.

LLMs could be derivatives of the books they memorize. Our results show that some LLMs memorize some works to such an extent that those copies can sometimes be generated as outputs (i.e., extraction). This matters for the dispute over whether LLMs themselves are

derivative works. A work is not a derivative work unless it is "substantially similar" in significant part to the original (Litchfield v. Spielberg). While prior debates over whether models can qualify as derivative works have been largely theoretical, we offer empirical evidence that is directly relevant to this issue. At least some LLMs may be derivative works of at least some books—because those LLMs have memorized a significant amount of protectable expression from those books. The inquiry does not end here; an LLM may be protected under copyright's **fair use** doctrine (Lemley & Casey, 2021). If an LLM memorizes content but only generates noninfringing outputs, that is a strong argument that how it makes use of that content is **transformative** and ought to be permissible fair use, rather than copyright infringement. But the fair use analysis of the LLM itself may be different from the analysis of the training dataset used for training (Lee et al., 2023b; Kadrey Judgment; Bartz Judgment).

Copyright law also treats intermediate copies differently from external-facing copies. An LLM is an intermediate work in many cases; it produces outputs, and those outputs are overwhelmingly not copyright infringements. But for companies that sell or release LLMs to others under open source licenses, the LLM itself is the product, and sometimes one that is being sold for commercial gain. That does not defeat the fair use argument, but it makes the analogy to fair use for training on copyrighted data less precise. The extent to which LLMs memorize is significantly influenced by choices during training, such as de-duplicating the training dataset (Lee et al., 2022). Policymakers may be less inclined to protect LLMs whose developers made training choices known to increase the risk of memorization—e.g., deliberate duplication of books (Appendix H.1.1).

It may sometimes be practical to generate infringing copies with an LLM. In Section 5, we use extraction of 50-token snippets for evidence of memorization, but no one would realistically extract an entire book this way. Long-form extraction requires other approaches, like the one in Section 6. This result for *Harry Potter* is not representative of most of the books we tested. Nevertheless, it is significant: for LLMs that exhibit high degrees of memorization of specific works, it may be possible to generate copies or derivatives of those works. Even though this may sometimes be possible, it may not always be easy to do, which may affect who courts view as directly liable for any output infringement (Lee et al., 2023b; Henderson et al., 2023). As a policy matter, the case for liability against the LLM developer is stronger if simple prompts generate infringing outputs (GEMA v. OpenAI). The case is weaker if a user employs more intensive strategies to try to generate infringing material (Appendix H.1.2).

Class actions are ill-suited for copyright claims about memorization. Because many of the pending cases are proceeding as **class actions**, plaintiffs will have to demonstrate that all book owners have sufficient common legal and fact issues, such that it makes sense for a court to certify the class and treat them all together (Fed. R. Civ. P. 23(a)(2)). As is clear from our results, that may be impossible; basic questions about whether a particular model actually incorporated any significant expression from any individual plaintiff's book cannot be generalized. The inclusion of a book in the training dataset does not mean that the resulting LLM memorized that book and whether, for that LLM, infringing output based on memorization is even possible. Instead, answering questions about the degree to which a specific LLM memorized a specific work requires experimental evidence. Some plaintiffs may be able to show copying, but others will not. Courts generally deny class certification in such circumstances (e.g., Wal-Mart Stores, Inc. v. Dukes) (Appendix H.1.3).

8 Conclusion

We develop (Section 3) and validate (Section 4) a sliding-window procedure to estimate the extent to which open-weight LLMs memorize different books, and find that the extent of books memorization is more significant than previously described. Memorization varies widely from model to model and from book to book (Section 5). With respect to our extraction procedure, most LLMs do not memorize most books, but this is not always the case. In extreme instances, an LLM may memorize a book so significantly that it is possible to extract the entire book near-verbatim, as we have shown with LLAMA 3.1 70B for *Harry Potter* (Section 6). These results complicate current disputes over copyright infringement, both by rejecting simplistic claims made by both sides about how LLMs work and by demonstrating that there is no single answer to the question of how much an LLM memorizes (Section 7).

Acknowledgments

This work was supported by a Stanford HAI Seed Research Grant.

We thank the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

We thank Franziska Boenisch, danah boyd, Nicholas Carlini, Alexandra Chouldechova, James Grimmelmann, Adam Dziedzic, Jamie Hayes, Matthew Jagielski, Kevin Klyman, Katherine Lee, Milad Nasr, Matthew Sag, Pamela Samuelson, Florian Tramèr, Danny Wilf-Townsend, and others for helpful discussion and comments on earlier versions of this project. We also thank Timothy B. Lee for his extremely clear explanation of our work (Lee, 2025), which helped us simplify the presentation of the math in this paper.

Ahmed Ahmed acknowledges generous support from a Knight-Hennessy Fellowship, an NSF Graduate Research Fellowship, and a Georgetown Foundation Research Grant.

A. Feder Cooper is an Assistant Professor of Computer Science at Yale University and a Postdoctoral Affiliate at Stanford University in Percy Liang's group in the Department of Computer Science and Daniel E. Ho's group at Stanford Law School. Until December 2025, Cooper was a full-time employee of Microsoft, working as a postdoctoral researcher within Microsoft Research. Nothing in this paper should be attributed to Microsoft.

Mark A. Lemley is a full-time professor at Stanford Law School. He also has a consulting practice as a partner at Lex Lumina LLP. As part of that practice he has represented AI companies in copyright litigation. Nothing in this paper should be attributed to those companies.

Percy Liang is a full-time professor at Stanford, and he is also a co-founder and consultant at Together AI, Simile AI, and Taiku AI; and also an advisor to Virtue AI and Humanitas AI. Nothing in this paper should be attributed to those companies.

Christopher De Sa is a full-time professor at Cornell, and he is also a consultant at Together AI, and an advisor at Sambanova Systems and Inception Labs. Nothing in this paper should be attributed to those companies.

Author contributions. AFC conceived of the project, wrote the code, designed and conducted the experiments, ran the analysis of the results, made the figures, and built the project website. MAL and ABC led the legal analysis, with support by AFC. AFC, MAL, and ABC drafted the manuscript. CDS provided technical feedback. CDS, MAL, PL, and DEH gave iterative feedback on visualizations. AC and AA helped compile books metadata and the supplementary material. PL, DEH, CDS, AA, and AC provided feedback on the manuscript. DEH provided the computing resources for the experiments. AG provided access to `Books3`.

Ethics Statement

Our work directly speaks to ongoing legal and ethical issues concerning training LLMs on copyrighted works, and has been cited in ongoing copyright litigation concerning generative AI.

We used the `Books3` LLM training dataset for our experiments, which is stored on a university computing cluster for research purposes only. This dataset is at the center of several copyright lawsuits (Kadrey v. Meta Platforms, Inc.); it originates from an ML researcher's torrent, and contains copies of hundreds of thousands of in-copyright books (Reisner, 2023; Lee et al., 2023b). We do not redistribute books that contain in-copyright material, but we redistribute the `Books3` copy of the CC-BY-SA-licensed book for which we run our analysis (*Down and Out in the Magic Kingdom* (Doctorow, 2003)).

We release our code repository publicly on GitHub, as well as a companion website for exploring book-text-free versions of our results. In the Appendix, we reference publicly available information pertaining to where ML researchers obtain `Books3` for research purposes.

References

17 U.S. Code § 503. Remedies for infringement: Impounding and disposition of infringing articles, December 2010. URL https://www.law.cornell.edu/uscode/text/17/503.

804 F.3d.

Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J. Hewett, Mojan Javaheripi, Piero Kauffmann, James R. Lee, Yin Tat Lee, Yuanzhi Li, Weishung Liu, Caio C. T. Mendes, Anh Nguyen, Eric Price, Gustavo de Rosa, Olli Saarikivi, Adil Salim, Shital Shah, Xin Wang, Rachel Ward, Yue Wu, Dingli Yu, Cyril Zhang, and Yi Zhang. Phi-4 Technical Report, 2024. URL https://arxiv.org/abs/2412.08905.

Chinua Achebe. *Things Fall Apart*. Anchor Books; Random House, Inc., 1958.

Douglas Adams. *The Hitchhiker's Guide to the Galaxy - Omnibus*. Pan Books; Pan Macmillan; William Heinemann Ltd, 1979.

Chimamanda Ngozi Adichie. *Americanah*. Alfred A. Knopf; Alfred A. Knopf Canada, 2013.

Shahad Al Rawi. *The Baghdad Clock*. Oneworld Publications; Dar al-Hikma, 2016.

amongglue/books3-subset-raw, 2025. URL https://huggingface.co/datasets/amongglue/books3-subset-raw.

Kelley Armstrong. *Industrial Magic*. Bantam Dell, 2004.

Isaac Asimov. *Fantastic Voyage*. Bantam Dell; Random House, Inc.; Bantam Books, 1966.

Isaac Asimov. *The Complete Robot*. Dream Letters Corporation; Dream Letters Corp., 1982.

Margaret Atwood. *The Handmaid's Tale*. Houghton Mifflin Harcourt Publishing Company, 1985.

Jane Austen. *Pride and Prejudice*. Penguin Books Ltd, 1813.

Author's Guild v. Google, Inc. 804 F.3d 202 (2d Cir. 2015).

David Baldacci. *The Christmas Train*. Warner Books, Inc.; Hachette Book Group, 2001.

James Baldwin. *Notes of a Native Son*. Beacon Press, 1955.

James Baldwin. *Another Country*. Michael Joseph; Penguin Books, 1962.

Julian Barnes. *The Lemon Table*. Vintage International, 2004.

Teodolinda Barolini. *Dante and the Origins of Italian Literary Culture*. Fordham University Press, 2006.

Bartz Judgment. Order on Fair Use, Bartz et al. v. Anthropic PBC. URL https://docs.justia.com/cases/federal/district-courts/california/candce/3:2024cv05417/434709/231. No. 3:23-cv-03417-VC (N.D. Cal. Jun. 25, 2025).

Mary Beard. *The Parthenon*. Profile Books Ltd, 2002.

Charles Beardsley. *Guam: Past and Present*. Charles E. Tuttle Company, Inc., 1964.

Samuel Beckett. *Waiting for Godot*. Grove Press; Grove/Atlantic, Inc., 1952.

Helen Benedict. *The Lonely Soldier*. Beacon Press, 2009.

Mary Berry. *Simple Cakes*. BBC Books; BBC Worldwide Ltd, 2006.

Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, et al. Pythia: A suite for analyzing large language models across training and scaling. In *International Conference on Machine Learning*, pp. 2397–2430. PMLR, 2023.

Emile Borel. La mecanique statique et l'irreversibilite. *J. Phys. Theor. Appl.*, 3(1):189–196, 1913. doi: 10.1051/jphystap:019130030018900. URL https://doi.org/10.1051/jphystap:019130030018900.

C.J. Box. *Paradise Valley*. Head of Zeus Ltd, 2017.

Ray Bradbury. *The Cat's Pajamas*. William Morrow; HarperCollins e-books, 2004.

Gillian Bradshaw. *London in Chains*. Severn House Publishers Ltd, 2009.

John Brockman. *My Einstein*. Pantheon Books; Random House of Canada Limited, 2006.

Dan Brown. *The Da Vinci Code*. Transworld Publishers; Bantam Press; Corgi, 2003.

Gareth Brown. *The Society of Unknowable Objects*. HarperCollins, 2025.

Niobia Bryant. *Live and Learn*. Dafina Books; Kensington Publishing Corp., 2007.

Mika Brzezinski. *Knowing Your Value*. Weinstein Books, 2011.

Albert Camus. *The Myth of Sisyphus*. Penguin Classics; Penguin Books Ltd, 1942.

CANBERT/pile_books3_text, 2025. URL https://huggingface.co/datasets/CANBERT/pile_books3_text.

Nicholas Carlini. What my privacy papers (don't) have to say about copyright and generative AI, 2025. URL https://nicholas.carlini.com/writing/2025/privacy-copyright-and-generative-models.html.

Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, pp. 2633–2650, 2021.

Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramer. Membership Inference Attacks From First Principles, 2022. URL https://arxiv.org/abs/2112.03570.

Nicholas Carlini, Daphne Ippolito, Matthew Jagielski, Katherine Lee, Florian Tramèr, and Chiyuan Zhang. Quantifying Memorization Across Neural Language Models. In *International Conference on Learning Representations*, 2023.

Lewis Carroll. *Alice's Adventures in Wonderland and Through the Looking-Glass and What Alice Found There*. Penguin Classics; Penguin Books Ltd, 1871.

Jeffrey A. Carver. *The Infinity Link*. Open Road Integrated Media, Inc., 1984.

Jacqueline Charlesworth. Generative AI's Illusory Case for Fair Use. *Vanderbilt Journal of Entertainment and Technology Law*, 27, 2025.

Chat GPT Is Eating the World, 2024. URL https://chatgptiseatingtheworld.com.

Agatha Christie. *Murder on the Orient Express*. HarperCollins e-books, 1933.

Agatha Christie. *And Then There Were None*. HarperCollins e-books, 1939.

CMG v. Anthropic. 3:23-cv-01092 (M.D. Tenn.).

Ta-Nehisi Coates. *The Beautiful Struggle*. Spiegel & Grau; Random House, Inc., 2008.

Ta-Nehisi Coates. *We Were Eight Years in Power*. One World; Random House; Penguin Random House LLC, 2017.

Ta-Nehisi Coates. *The Water Dancer*. One World; Random House; Penguin Random House LLC, 2019.

Jean Cocteau. *The Infernal Machine and Other Plays*. New Directions; New Directions Publishing Corporation, 1963.

Paulo Coelho. *The Alchemist*. HarperCollins Publishers, 1988.

Jon Cogburn and Mark Silcox. *Dungeons and Dragons and Philosophy*. Open Court Publishing Company, 2012.

Annie Cohen-Solal. *Mark Rothko*. Yale University Press, 2013.

Jasmine Collins, Jascha Sohl-Dickstein, and David Sussillo. Capacity and Trainability in Recurrent Neural Networks, 2017. URL https://arxiv.org/abs/1611.09913.

Suzanne Collins. *The Hunger Games*. Scholastic Children's Books; Scholastic Ltd, 2008.

Glen Cook. *The Dragon Never Sleeps*. Night Shade Books, 1988.

A. Feder Cooper and James Grimmelmann. The Files are in the Computer: Copyright, Memorization, and Generative AI. *arXiv preprint arXiv:2404.12590*, 2024.

A. Feder Cooper, Katherine Lee, James Grimmelmann, Daphne Ippolito, Christopher Callison-Burch, Christopher A. Choquette-Choo, Niloofar Mireshghallah, Miles Brundage, David Mimno, Madiha Zahrah Choksi, Jack M. Balkin, Nicholas Carlini, Christopher De Sa, Jonathan Frankle, Deep Ganguli, Bryant Gipson, Andres Guadamuz, Swee Leng Harris, Abigail Z. Jacobs, Elizabeth Joh, Gautam Kamath, Mark Lemley, Cass Matthews, Christine McLeavey, Corynne McSherry, Milad Nasr, Paul Ohm, Adam Roberts, Tom Rubin, Pamela Samuelson, Ludwig Schubert, Kristen Vaccaro, Luis Villa, Felix Wu, and Elana Zeide. Report of the 1st Workshop on Generative AI and Law. *arXiv preprint arXiv:2311.06477*, 2023.

A. Feder Cooper, Christopher A. Choquette-Choo, Miranda Bogen, Matthew Jagielski, Katja Filippova, Ken Ziyu Liu, Alexandra Chouldechova, Jamie Hayes, Yangsibo Huang, Niloofar Mireshghallah, Ilia Shumailov, Eleni Triantafillou, Peter Kairouz, Nicole Mitchell, Percy Liang, Daniel E. Ho, Yejin Choi, Sanmi Koyejo, Fernando Delgado, James Grimmelmann, Vitaly Shmatikov, Christopher De Sa, Solon Barocas, Amy Cyphert, Mark Lemley, danah boyd, Jennifer Wortman Vaughan, Miles Brundage, David Bau, Seth Neel, Abigail Z. Jacobs, Andreas Terzis, Hanna Wallach, Nicolas Papernot, and Katherine Lee. Machine Unlearning Doesn't Do What You Think: Lessons for Generative AI Policy, Research, and Practice. *arXiv preprint arXiv:2412.06966*, 2024.

A. Feder Cooper, Aaron Gokaslan, Amy B. Cyphert, Christopher De Sa, Mark A. Lemley, Daniel E. Ho, and Percy Liang. Extracting memorized pieces of (copyrighted) books from open-weight language models. *arXiv preprint arXiv:2505.12546*, 2025.

A. Feder Cooper, Marika Swanberg, Jamie Hayes, Lea Duesterwald, Christopher De Sa, Daniel E. Ho, Mark A. Lemley, and Percy Liang. Extractable Memorization From First Principles. *arXiv preprint arXiv:2607.12649*, 2026.

CoStar Grp., Inc. v. LoopNet, Inc. 373 F.3d 544 (4th Cir. 2004).

Stephen R. Covey. *The 7 Habits of Highly Effective People*. RosettaBooks LLC, 1989.

Tyler Cowen. Toothpick Producers Violate NYT Copyright. *Marignal Revolution*, December 2023. URL https://marginalrevolution.com/marginalrevolution/2023/12/toothpick-producers-violate-nyt-copyright.html.

David Crabb. *Bad Kid*. HarperCollins e-books; HarperCollins Publishers Inc., 2015.

Robert Crais. *Lullaby Town*. Bantam Books; Random House, Inc., 1992.

Michael Crichton. *Jurassic Park*. Ballantine Books; The Random House Publishing Group, 1990.

Michael Cunningham. *The Hours*. Picador; Farrar, Straus and Giroux, 1998.

Amy B. Cyphert. Generative AI, Plagiarism, and Copyright Infringement in Legal Documents. *Minnesota Journal of Law, Science & Technology*, 25, 2024.

Józef Czapski. *Inhuman Land*. New York Review Books; The New York Review of Books, 1949.

Roald Dahl. *James and the Giant Peach*. Puffin Books, 1961.

Roald Dahl. *Charlie and the Chocolate Factory*. Puffin Books; Penguin Books Ltd, 1964.

DeepSeek-AI et al. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism, 2024. URL https://arxiv.org/abs/2401.02954.

Nicholas Diakopoulos. *Automating the News*. Harvard University Press, 2019.

Joan Didion. *The White Album*. Zola Books; Simon & Schuster, 1979.

Cory Doctorow. *Down and Out in the Magic Kingdom*. Tor; Tom Doherty Associates, LLC, 2003.

Carol Ann Duffy. *The World's Wife*. Picador; Pan Macmillan, 1999.

Junot Díaz. *Drown*. Riverhead Books; The Berkley Publishing Group, 1995.

Junot Díaz. *The Brief Wondrous Life of Oscar Wao*. Riverhead Books; Penguin Group (USA) Inc., 2007.

Junot Díaz. *This Is How You Lose Her*. Riverhead Books; Penguin Group (USA) Inc., 2012.

Jennifer Egan. *A Visit from the Goon Squad*. Alfred A. Knopf; Random House of Canada Limited, 2010.

EleutherAI/the_pile_deduplicated, 2025. URL https://huggingface.co/datasets/EleutherAI/the_pile_deduplicated.

Ralph Ellison. *Invisible Man*. Penguin Books; Penguin Random House, 1952.

Chelsea Fairless and Lauren Garroni. *We Should All Be Mirandas*. Houghton Mifflin Harcourt Publishing Company, 2019.

Cathleen Falsani. *The Dude Abides*. Zondervan, 2009.

Christopher Farnsworth. *The President's Vampire*. G. P. Putnam's Sons; Penguin Books Ltd, 2011.

Fed. R. Civ. P. 23(a)(2). (requiring questions of law or fact common to the class before it can be certified).

F. Scott Fitzgerald. *The Great Gatsby*. Harper Perennial Classics; HarperCollins Publishers (Canada) Ltd., 1925.

Gillian Flynn. *Gone Girl*. Crown Publishers, 2012.

Norman Friedman. *British Destroyers*. Seaforth Publishing; Pen & Sword Books Ltd, 2009.

Barbara Fudurich, Joan Hansen, Caroline Linscott, and Geri Medway. *Florals & Botanicals*. Walter Foster Publishing, Inc., 2008.

Neil Gaiman and Terry Pratchett. *Good Omens*. HarperCollins Publishers, 1990.

Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *arXiv preprint arXiv:2101.00027*, 2020.

GEMA v. OpenAI. Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte. 42 O 14139/24.

Gemini Team, Nikolay Savinov, Denis Teplyashin, Dmitry Lepikhin, Timothy Lillicrap, Jean-baptiste Alayrac, Radu Soricut, Angeliki Lazaridou, Orhan Firat, Julian Schrittwieser, et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

Melanie Gideon. *The Slippery Year*. Alfred A. Knopf; Random House of Canada Limited, 2009.

Stefanie Girard. *Sweater Surgery*. Quarry Books; Quayside Publishing Group, 2010.

Malcolm Gladwell. *Blink*. Penguin Books; Allen Lane; Little, Brown and Company, 2005.

Richard Glover. *The Land Before Avocado*. HarperCollins Publishers Australia Pty Limited, 2018.

Christopher Golden. *Dead Ringers*. St Martin's Press, 2015.

Christopher Golden. *Ararat*. St. Martin's Press, 2017.

William Golding. *Lord of the Flies*. Penguin Books; Penguin Random House LLC, 1954.

Terry Goodkind. *Wizard's First Rule*. RosettaBooks LLC; Tor Books, 1994.

Martin Goodman. *Rome and Jerusalem*. Alfred A. Knopf; Allen Lane, 2007.

Michael P. Goodyear. Artificial Infringement. In *Proceedings of the 2025 Symposium on Computer Science and Law*, CSLAW '25, pp. 26–38, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714214. doi: 10.1145/3709025.3712208. URL https://doi.org/10.1145/3709025.3712208.

Google Cloud Blog. BFloat16: The Secret to High Performance on Cloud TPUs, 2019. URL https://cloud.google.com/blog/products/ai-machine-learning/bfloat16-the-secret-to-high-performance-on-cloud-tpus. "the dynamic range of bfloat16 is identical to that of FP32".

Aaron Grattafiori et al. The Llama 3 Herd of Models, 2024. URL https://arxiv.org/abs/2407.21783.

John Green. *The Fault in Our Stars*. Dutton Books; Penguin Group (USA) Inc., 2012.

Graham Greene. *The Third Man*. Vintage Digital; Vintage Press, 1950.

Andrew Sean Greer. *The Confessions of Max Tivoli*. Picador; Farrar, Straus and Giroux, 2004.

John Grisham. *The Fugitive*. Dutton Children's Books; Penguin Group (USA) LLC, 2015.

Mark Haddon. *The Curious Incident of the Dog in the Night-Time*. Vintage Books; Random House of Canada Limited, 2003.

Sue Halpern. *Migrations to Solitude*. Pantheon Books; Random House of Canada Limited; Vintage Books, 1992.

Tom Hanks. *Uncommon Type*. Alfred A. Knopf, 2017.

Thor Hanson. *Buzz*. Icon Books Ltd; Basic Books, 2018.

Harry Potter Wiki. Mistakes in the Harry Potter books, 2025. URL https://harrypotter.fandom.com/wiki/Mistakes_in_the_Harry_Potter_books.

Elizabeth Haydon. *Requiem for the Sun*. Tor Books; Tom Doherty Associates, LLC, 2002.

Jamie Hayes, Ilia Shumailov, Christopher A. Choquette-Choo, Matthew Jagielski, George Kaissis, Katherine Lee, Milad Nasr, Sahra Ghalebikesabi, Niloofar Mireshghallah, Meenatchi Sundaram Mutu Selva Annamalai, Igor Shilov, Matthieu Meeus, Yves-Alexandre de Montjoye, Franziska Boenisch, Adam Dziedzic, and A. Feder Cooper. Strong Membership Inference Attacks on Massive Datasets and (Moderately) Large Language Models. *arXiv preprint arXiv:2505.18773*, 2025a.

Jamie Hayes, Marika Swanberg, Harsh Chaudhari, Itay Yona, Ilia Shumailov, Milad Nasr, Christopher A. Choquette-Choo, Katherine Lee, and A. Feder Cooper. Measuring memorization in language models via probabilistic extraction. In Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 9266–9291, Albuquerque, New Mexico, April 2025b. Association for Computational Linguistics. ISBN 979-8-89176-189-6. URL https://aclanthology.org/2025.naacl-long.469/.

Joseph Heller. *Catch-22*. Simon & Schuster Paperbacks; Simon & Schuster, 1961.

Ernest Hemingway. *The Old Man and the Sea*. Harper Perennial Classics; HarperCollins Publishers (Canada) Ltd., 1952.

Peter Henderson, Xuechen Li, Dan Jurafsky, Tatsunori Hashimoto, Mark A. Lemley, and Percy Liang. Foundation Models and Fair Use, 2023. URL https://arxiv.org/abs/2303.15715.

David Hendy. *Life on Air*. Oxford University Press; Oxford University Press Inc., 2007.

Emily Henry. *Great Big Beautiful Life*. Berkley, 2025.

Luke Hersheson. *Great Hair Days*. Ebury Press; Penguin Random House, 2018.

Matt Hill. *Graft*. Angry Robot; Watkins Media Ltd, 2016.

S. E. Hinton. *The Outsiders*. Penguin Books; The Viking Press; Puffin Books, 1967.

Tanya Huff. *The Second Summoning*. DAW Books, 2001.

Hugging Face. Quantization Concepts, 2025. URL https://huggingface.co/docs/optimum/v1.7.1/concept_guides/quantization. "bfloat16, accumulation data type float32".

Langston Hughes. *Selected Poems of Langston Hughes*. Vintage Books; Random House of Canada Limited, 1959.

David Henry Hwang. *M. Butterfly*. Plume; Penguin Group (USA) Inc., 1988.

Mike Isaac. How A.I. Made Mark Zuckerberg Popular Again in Silicon Valley. *The New York Times*, May 2024. URL https://www.nytimes.com/2024/05/29/technology/mark-zuckerberg-meta-ai.html.

Allen Jackson. *Building and Operating a Realistic Model Railway*. The Crowood Press Ltd, 2016.

Betty Jacobs. *All the Onions*. Storey Publishing, LLC, 1993.

E. L. James. *Fifty Shades of Grey*. The Writer's Coffee Shop Publishing House; Vintage Books; Random House of Canada Limited, 2011.

N. K. Jemisin. *The Stone Sky*. Orbit, 2017.

James Joyce. *Ulysses*. Penguin Books Ltd, 1922.

Richard Kadrey. *Sandman Slim*. HarperCollins e-books; HarperCollins Publishers Inc., 2009.

Kadrey Judgment. Order Denying the Plaintiffs' Motion for Partial Summary Judgment and Granting Meta's Cross-Motion for Partial Summary Judgment, Kadrey et al. v. Meta Platforms, Inc. URL https://law.justia.com/cases/federal/district-courts/california/candce/3:2023cv03417/415175/598/. No. 3:23-cv-03417-VC (N.D. Cal. Jun. 25, 2025).

Kadrey v. Meta. URL https://www.courthousenews.com/wp-content/uploads/2025/02/kadrey-vs-meta-third-amended-complaint.pdf. No. 3:23-cv-03417-VC (Third Amended Consolidated Complaint).

Kadrey v. Meta Platforms, Inc. No. 3:23-cv-03417 (N.D. Cal. July 7, 2023).

Anthony Kaldellis. *Ethnography after Antiquity*. University of Pennsylvania Press, 2013.

Antonia Karamolegkou, Jiaang Li, Li Zhou, and Anders Søgaard. Copyright Violations and Large Language Models, 2023. URL https://arxiv.org/abs/2310.13771.

Matthew Klam. *Who Is Rich?* Random House; Penguin Random House LLC, 2017.

Dean Koontz. *The Servants of Twilight*. The Berkley Publishing Group; Penguin Group (USA) Inc., 1984.

Aihan Kuhn. *Tai Chi for Depression*. YMAA Publication Center, Inc., 2017.

Josh Kun. *The Tide Was Always High*. University of California Press, 2017.

Jean Kwok. *Girl in Translation*. Riverhead Books; Penguin Group (USA) Inc., 2010.

Eduardo Lago. *Call Me Brooklyn*. Dalkey Archive Press, 2006.

Erik Larson. *Dead Wake*. Crown Publishers, 2015.

Stieg Larsson. *The Girl with the Dragon Tattoo*. Alfred A. Knopf, 2005.

Ursula K. Le Guin. *The Daughter of Odren*. Houghton Mifflin Harcourt Publishing Company, 2014.

Dave Lee. Facebook denies targeting insecure users. *BBC*, May 2017. URL https://www.bbc.com/news/technology-39775782.

Katherine Lee, Daphne Ippolito, Andrew Nystrom, Chiyuan Zhang, Douglas Eck, Chris Callison-Burch, and Nicholas Carlini. Deduplicating Training Data Makes Language Models Better. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, volume 1, pp. 8424–8445, 2022.

Katherine Lee, A. Feder Cooper, James Grimmelmann, and Daphne Ippolito. AI and Law: The Next Generation. *SSRN*, 2023a. http://dx.doi.org/10.2139/ssrn.4580739.

Katherine Lee, A. Feder Cooper, and James Grimmelmann. Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain. *arXiv preprint arXiv:2309.08133*, 2023b.

Katherine Lee, A. Feder Cooper, and James Grimmelmann. Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain (The Short Version). In *Proceedings of the Symposium on Computer Science and Law*, CSLAW '24, pp. 48–63, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703331. doi: 10.1145/3614407.3643696. URL https://doi.org/10.1145/3614407.3643696.

Timothy B. Lee. Study: Meta AI model can reproduce almost half of Harry Potter book. *Ars Technica*, June 2025. URL https://arstechnica.com/ai/2025/06/study-metas-llama-3-1-can-recall-42-percent-of-the-first-harry-potter-book/.

Mark Lemley and Bryan Casey. Fair Learning. *Texas Law Review*, 99:743, 2021.

Madeleine L'Engle. *A Wrinkle in Time*. Puffin Books; Penguin Books Ltd, 1962.

C. S. Lewis. *The Chronicles of Narnia*. Penguin Random House, 1950.

Laura Lippman. *After I'm Gone*. HarperCollins e-books, 2014.

Laura Lippman. *Sunburn*. Faber & Faber Ltd; William Morrow, 2018.

David Liss. *Marvel's Spider-Man: Hostile Takeover*. Titan Books; Titan Publishing Group Ltd, 2018.

Litchfield v. Spielberg. 736 F.2d 1352, 1357 (9th Cir. 1984). ("a work is not derivative unless it has been substantially copied from the prior work.").

Lois Lowry. *Anastasia on Her Own*. Houghton Mifflin Books for Children; Houghton Mifflin Harcourt Publishing Company, 1985.

Iris Ma, Ian Domingo, Alberto Krone-Martins, Pierre Baldi, and Cristina V. Lopes. Memorization: A Close Look at Books, 2025. URL https://arxiv.org/abs/2504.12549.

Margaret MacMillan. *Nixon in China*. Penguin Group (Canada), 2006.

Sadegh Mahdavi, Renjie Liao, and Christos Thrampoulidis. Memorization Capacity of Multi-Head Attention in Transformers, 2024. URL https://arxiv.org/abs/2306.02010.

Scott Mariani. *The Doomsday Prophecy*. Avon; HarperCollins Publishers Ltd., 2009.

George R. R. Martin. *A Game of Thrones*. Bantam Books, 1996.

Keith Heyer Meldahl. *Rough-Hewn Land*. University of California Press; University of California Press, Ltd., 2011.

Meta Newsroom. Celebrating 1 Billion Downloads of Llama, March 2025. URL https://about.fb.com/news/2025/03/celebrating-1-billion-downloads-llama/.

Stephenie Meyer. *Twilight*. Little, Brown and Company; Hachette Book Group, 2005.

Courtney Milan. *The Duchess War*. Courtney Milan (self-published), 2012.

John Stuart Mill. *On Liberty*. Penguin Books, 1859.

Alex Miller. *Coal Creek*. Allen & Unwin, 2013.

A. A. Milne. *Winnie the Pooh*. Dutton Children's Books; Penguin Group (USA) Inc., 1926.

Stephanie Mlot. Zuckerberg Vows to Protect Free Speech After Paris Attack. *PCMag*, January 2015. URL https://www.pcmag.com/news/zuckerberg-vows-to-protect-free-speech-after-paris-attack.

Colten Moore and Keith O'Brien. *Catching the Sky*. 37 INK/Atria Books; Simon & Schuster, Inc., 2016.

Michelle Moran. *The Heretic Queen*. Crown Publishers; Random House, Inc., 2008.

Laura Morelli. *The Gondola Maker*. Laura Morelli; Smashwords, 2014.

Mary McGarry Morris. *Songs in Ordinary Time*. Open Road Integrated Media, Inc., 1995.

Toni Morrison. *Beloved*. Vintage Books; Random House of Canada Limited; Alfred A. Knopf, 1987a.

Toni Morrison. Beloved, 1987b. URL https://www.goodreads.com/quotes/386603-in-this-here-place-we-flesh-flesh-that-weeps-laughs. Part One, Chapter 9.

Toni Morrison. Beloved, 1987c. URL https://www.goodreads.com/quotes/11207714-leave-before-sethe-could-make-her-realize-that-worse-than. Part One, Chapter 25.

Toni Morrison. Beloved, 1987d. URL https://www.goodreads.com/quotes/578852-i-used-to-think-to-think-it-was-my-rememory. Part One, Chapter 9.

Haruki Murakami. *Norwegian Wood*. Vintage; The Harvill Press; Kodansha Ltd, 1987.

Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Eric Wallace, Florian Tramèr, and Katherine Lee. Scalable Extraction of Training Data from (Production) Language Models. *arXiv preprint arXiv:2311.17035*, 2023.

Milad Nasr, Javier Rando, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Florian Tramèr, and Katherine Lee. Scalable Extraction of Training Data from Aligned, Production Language Models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=vjel3nWP2a.

Desiree Nielsen. *Eat More Plants*. Penguin Canada; Penguin Random House Canada Limited, 2019.

Michael Northrop. *Polaris*. Scholastic Press; Scholastic Inc., 2017.

Barack Obama. Keynote Address at the 2004 Democratic National Convention, 2004. URL https://www.presidency.ucsb.edu/documents/keynote-address-the-2004-democratic-national-convention.

James J. O'Donnell. *Pagans*. HarperCollins e-books, 2015.

Yoko Ogawa. *The Memory Police*. Pantheon Books, 1994.

Robin Oliveira. *Winter Sisters*. Viking; Penguin Random House LLC, 2018.

OpenAI. OpenAI and journalism, January 2024. URL https://openai.com/blog/openai-and-journalism.

George Orwell. *Nineteen Eighty-Four*. Penguin Books Ltd; Martin Secker & Warburg Ltd, 1949.

Meghan L. O'Sullivan. *Windfall*. Simon & Schuster, 2017.

Chuck Palahniuk. *Fight Club*. W. W. Norton & Company, Inc.; W. W. Norton & Company Ltd., 1996.

Charlie Papazian. *The Complete Joy of Homebrewing*. HarperCollins Publishers Inc., 1984.

Paul Park. *The Cult of Loving Kindness*. ElectricStory.com, Inc.; William Morrow, 1991.

James Patterson. *Along Came a Spider*. Little, Brown and Company; Hachette Book Group, 1993.

François Payard and Anne E. McBride. *Payard Cookies*. Houghton Mifflin Harcourt, 2015.

Tom Piazza. *Why New Orleans Matters*. HarperCollins Publishers; Harper Perennial; HarperCollins e-books, 2005.

Steven Pinker. *Enlightenment Now*. Viking; Penguin Random House LLC, 2018.

Michael E. Porter. *Competitive Strategy*. The Free Press; Simon & Schuster Inc., 1980.

USVSN Sai Prashanth, Alvin Deng, Kyle O'Brien, Jyothir S V au2, Mohammad Aflah Khan, Jaydeep Borkar, Christopher A. Choquette-Choo, Jacob Ray Fuehne, Stella Biderman, Tracy Ke, Katherine Lee, and Naomi Saphra. Recite, Reconstruct, Recollect: Memorization in LMs as a Multifaceted Phenomenon, 2024. URL https://arxiv.org/abs/2406.17746.

Terry Pratchett. *Night Watch*. HarperCollins Publishers, 2002.

Philip Pullman. *The Subtle Knife*. Alfred A. Knopf, Inc.; Scholastic Children's Books, 1997.

Jacques Pépin. *Essential Pepin Desserts*. Houghton Mifflin Harcourt, 2011.

Qwen, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 Technical Report, 2025. URL https://arxiv.org/abs/2412.15115.

Matthew Rampley. *The Seductions of Darwin*. The Pennsylvania State University Press, 2017.

Alex Reisner. These 183,000 Books Are Fueling the Biggest Fight in Publishing and Tech. *The Atlantic*, September 2023. URL https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/.

Rachel Naomi Remen. *Kitchen Table Wisdom*. Riverhead Books; Penguin Group (USA) Inc., 1996.

Jessi Roberts and Bret Witter. *Backroads Boss Lady*. Grand Central Publishing; Hachette Digital, 2019.

Marie-Sabine Roger. *Soft in the Head*. Pushkin Press, 2008.

Ramzi Rouighi. *The Making of a Mediterranean Emirate*. University of Pennsylvania Press, 2011.

J. K. Rowling. *Harry Potter and the Sorcerer's Stone*. Pottermore Limited; Arthur A. Levine Books, 1997.

J. K. Rowling. *Harry Potter and the Chamber of Secrets*. Pottermore Limited; Arthur A. Levine Books, 1998.

J. K. Rowling. *Harry Potter and the Goblet of Fire*. Pottermore Limited; Arthur A. Levine Books; Scholastic Inc., 2000.

J. K. Rowling. *Harry Potter and the Deathly Hallows*. Pottermore Limited; Arthur A. Levine Books; Scholastic Inc., 2007.

Dan Rubinstein. *Born to Walk*. ECW Press, 2015.

Marc Ruskin. *The Pretender*. Thomas Dunne Books; St. Martin's Press, 2017.

Harvey Sachs. *Toscanini*. Liveright Publishing Corporation; W. W. Norton & Company, Inc., 1978.

Matthew Sag. Copyright Safety for Generative AI. *Houston Law Review*, 61, 2024a.

Matthew Sag. Fairness and Fair Use in Generative AI. *Fordham Law Review*, 92, 2024b.

Carl Sagan. *Cosmos*. Ballantine Books; Random House of Canada Limited, 1980.

Bhisham Sahni. *Middle India*. Penguin Books India Pvt. Ltd, 2001.

J. D. Salinger. *The Catcher in the Rye*. Little, Brown and Company, 1951.

Pamela Samuelson. Generative AI meets copyright. *Science*, 381(6654):158–161, 2023.

Sheryl Sandberg. *Lean in*. Alfred A. Knopf; Random House of Canada Limited, 2013.

SaylorTwift/the_pile_books3_minus_gutenberg, 2025. URL https://huggingface.co/datasets/SaylorTwift/the_pile_books3_minus_gutenberg.

Avi Schwarzschild, Zhili Feng, Pratyush Maini, Zachary C Lipton, and J Zico Kolter. Rethinking llm memorization through the lens of adversarial compression. *arXiv preprint arXiv:2404.15146*, 2024.

Sega Enters. Ltd. v. Accolade, Inc. 977 F.2d 1510 (9th Cir. 1992), as amended (Jan. 6, 1993).

Aarti Shahani. Facebook Wants Great Power, But What About Responsibility? *NPR*, February 2017. URL https://www.npr.org/sections/alltechconsidered/2017/02/19/516094134/facebook-wants-great-power-but-what-about-responsibility.

Sanjeev Sharma. *The DevOps Adoption Playbook*. John Wiley & Sons, Inc., 2017.

Mary Shelley. *Frankenstein*. Wordsworth Classics, 1818.

Lynn Sherr. *Sally Ride: America's First Woman in Space*. Simon & Schuster, 2014.

Lionel Shriver. *A Perfectly Good Family*. HarperCollins Publishers Inc., 1996.

Sarah Silverman. *The Bedwetter*. HarperCollins e-books, 2010.

Larry Sloman. *On the Road with Bob Dylan*. Three Rivers Press; Bantam Books, 1978.

Alexander Gordon Smith. *The Night Children*. Macmillan, 2008.

Zadie Smith. *White Teeth*. Vintage International, 2000.

Rachel Louise Snyder. *No Visible Bruises*. Bloomsbury Publishing Inc., 2019.

Benjamin Sobel. Elements of Style: Copyright, Similarity, and Generative AI. *Harvard Journal of Law & Technology*, 38, 2024.

Sony Computer Ent., Inc. v. Connectix Corp. 203 F.3d 596 (9th Cir. 2000).

John Steinbeck. *The Grapes of Wrath*. Penguin Books; The Viking Press Inc.; William Heinemann Ltd, 1939.

T. J. Stiles. *Jesse James*. Vintage Books; Random House of Canada Limited; Alfred A. Knopf, 2002.

R. L. Stine. *Zombie Halloween*. Scholastic Inc., 2014.

David Stubbs. *Fear of Music*. Zero Books; John Hunt Publishing Ltd., 2009.

Jennifer Swanson. *Pearl Harbor*. Scholastic Inc., 2018.

Earl Swift. *Chesapeake Requiem*. HarperCollins Publishers, 2018.

Donna Tartt. *The Goldfinch*. Little, Brown and Company; Hachette Book Group, 2013.

Gemma Team et al. Gemma 2: Improving Open Language Models at a Practical Size, 2024. URL https://arxiv.org/abs/2408.00118.

Lysa TerKeurst. *Unglued*. Zondervan, 2012.

Lysa TerKeurst. *Embraced*. Thomas Nelson, 2018.

The Bible (King James Version). Deuteronomy 15:12-15. URL https://www.kingjamesbibleonline.org/Deuteronomy-15-12/.

The Simpsons: Last Exit to Springfield (Fox television broadcast Mar. 11, 1993), 1993. ("This is a thousand monkeys working at a thousand typewriters. Soon they'll have written the greatest novel known to man . . . [l]et's see. 'It was the best of times, it was the blurst of times?' You stupid monkey!").

Lynn Thomson. *Birding with Yeats*. House of Anansi Press Inc., 2014.

J. R. R. Tolkien. *The Hobbit*. Houghton Mifflin Harcourt, 1937.

J. R. R. Tolkien. *The Fellowship of the Ring*. Houghton Mifflin Harcourt; George Allen & Unwin, 1954.

J. R. R. Tolkien. *Tree and Leaf*. HarperCollins Publishers, 1964.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and Efficient Foundation Language Models, 2023a. URL https://arxiv.org/abs/2302.13971.

Hugo Touvron et al. Llama 2: Open Foundation and Fine-Tuned Chat Models, 2023b. URL https://arxiv.org/abs/2307.09288.

Corinne Trang. *Noodles Every Day*. Chronicle Books LLC, 2009.

George R. Tyler. *Billionaire Democracy*. BenBella Books, Inc., 2018.

United States Copyright Office. Copyright and Artificial Intelligence (Part 2: Copyrightability), 2025. URL https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf.

Al J. Venter. *Portugal's Guerrilla Wars in Africa*. Helion & Company Limited, 2013.

Kurt Vonnegut. *Slaughterhouse-Five*. RosettaBooks LLC, 1969.

Ocean Vuong. *The Emperor of Gladness*. Penguin Press, 2025.

Wal-Mart Stores, Inc. v. Dukes. 564 U.S. 338, 349–50 (2011). (holding that class certification was inappropriate because commonality requires that class members be capable of showing that they have suffered the same injury).

Paul Waldau. *Animal Rights*. Oxford University Press, Inc., 2011.

Jesmyn Ward. *Men We Reaped*. Bloomsbury USA, 2013.

E. B. White. *Charlotte's Web*. HarperCollins e-books; HarperCollins Publishers Inc., 1952.

Daniel Wilf-Townsend. The Deletion Remedy. *North Carolina Law Review*, 103, 2024. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4933011. Forthcoming 2025.

Marianne Williamson. *A Return to Love*. HarperCollins Publishers, 1992.

Julia Carrie Wong. SpaceX rocket explosion: Mark Zuckerberg laments loss of Internet.org satellite. *The Guardian*, September 2016. URL https://www.theguardian.com/science/2016/sep/01/mark-zuckerberg-spacex-explosion-africa-internet.

Jacqueline Woodson. *Brown Girl Dreaming*. Nancy Paulsen Books; Penguin Group (USA) LLC, 2014.

Jacqueline Woodson. *Another Brooklyn*. HarperCollins e-books, 2016.

Sarah Wynn-Williams. *Careless People: A Cautionary Tale of Power, Greed, and Lost Idealism*. Macmillan, 2025.

Malcolm X. Who Taught You To Hate Yourself?, 1962. URL https://genius.com/Malcolm-x-who-taught-you-to-hate-yourself-annotated.

Hanya Yanagihara. *A Little Life*. Doubleday; Random House of Canada Limited, 2015.

Yuji Yoshimura and Giovanna M. Halford. *The Art of Bonsai*. Tuttle Publishing; Charles E. Tuttle Publishing Co, Inc., 1996.

Chiyuan Zhang, Daphne Ippolito, Katherine Lee, Matthew Jagielski, Florian Tramèr, and Nicholas Carlini. Counterfactual memorization in neural language models. *Advances in Neural Information Processing Systems*, 36:39321–39362, 2023.

Howard Zinn. *A People's History of the United States*. HarperCollins Publishers; Harper Perennial Modern Classics, 1980.

Jonathan Zittrain. *The Future of the Internet and How to Stop It*. Yale University Press, 2008.

Markus Zusak. *The Book Thief*. Alfred A. Knopf; Random House of Canada Limited; Picador; Pan Macmillan Pty Ltd., 2005.

Appendix Table of Contents

A Additional notes on memorization and probabilistic extraction 30
- **A.1** Additional notes on the "careless people" example 30
- **A.2** Metrics 30
- **A.3** Computing sequence probabilities in one forward pass 32

B Background on memorization, extraction risk, and U.S. copyright 33

C Testing our measurement pipeline 35

D Experiments on extraction rates 36
- **D.1** Setup 36
- **D.2** Experimental procedure 37
- **D.3** High-level takeaways 37

E Extended results for sliding-window experiments 39
- **E.1** Setup 39
- **E.2** Sliding-window probabilistic extraction procedure 50
- **E.3** Book-specific results 52
  - **E.3.1** *Things Fall Apart*, Achebe 53
  - **E.3.2** *The Hitchhiker's Guide to the Galaxy - Omnibus*, Adams 53
  - **E.3.3** *Americanah*, Adichie 54
  - **E.3.4** *The Baghdad Clock*, Al Rawi 54
  - **E.3.5** *Industrial Magic*, Armstrong 55
  - **E.3.6** *Fantastic Voyage*, Asimov 55
  - **E.3.7** *The Complete Robot*, Asimov 56
  - **E.3.8** *The Handmaid's Tale*, Atwood 56
  - **E.3.9** *Pride and Prejudice*, Austen 57
  - **E.3.10** *The Christmas Train*, Baldacci 57
  - **E.3.11** *Notes of a Native Son*, Baldwin 58
  - **E.3.12** *Another Country*, Baldwin 58
  - **E.3.13** *The Lemon Table*, Barnes 59
  - **E.3.14** *Dante and the Origins of Italian Literary Culture*, Barolini 59
  - **E.3.15** *The Parthenon*, Beard 60
  - **E.3.16** *Guam: Past and Present*, Beardsley 60
  - **E.3.17** *Waiting for Godot*, Beckett 61
  - **E.3.18** *The Lonely Soldier*, Benedict 61
  - **E.3.19** *Simple Cakes*, Berry 62
  - **E.3.20** *Paradise Valley*, Box 62
  - **E.3.21** *The Cat's Pajamas*, Bradbury 63
  - **E.3.22** *London in Chains*, Bradshaw 63
  - **E.3.23** *My Einstein*, Brockman 64
  - **E.3.24** *The Da Vinci Code*, Brown 64
  - **E.3.25** *Live and Learn*, Bryant 65
  - **E.3.26** *Knowing Your Value*, Brzezinski 65
  - **E.3.27** *The Myth of Sisyphus*, Camus 66
  - **E.3.28** *Alice's Adventures in Wonderland*, Carroll 66
  - **E.3.29** *The Infinity Link*, Carver 67
  - **E.3.30** *Murder on the Orient Express*, Christie 67
  - **E.3.31** *And Then There Were None*, Christie 68
  - **E.3.32** *The Beautiful Struggle*, Coates 68
  - **E.3.33** *We Were Eight Years in Power*, Coates 69
  - **E.3.34** *The Water Dancer*, Coates 69
  - **E.3.35** *The Infernal Machine*, Cocteau 70
  - **E.3.36** *The Alchemist*, Coelho 70

E.3.37 *Dungeons and Dragons and Philosophy*, Cogburn & Silcox 71
E.3.38 *Mark Rothko*, Cohen-Solal 71
E.3.39 *The Hunger Games*, Collins 72
E.3.40 *The Dragon Never Sleeps*, Cook 72
E.3.41 *The 7 Habits of Highly Effective People*, Covey 73
E.3.42 *Bad Kid*, Crabb 73
E.3.43 *Lullaby Town*, Crais 74
E.3.44 *Jurassic Park*, Crichton 74
E.3.45 *The Hours*, Cunningham 75
E.3.46 *Inhuman Land*, Czapski 75
E.3.47 *Charlie and the Chocolate Factory*, Dahl 76
E.3.48 *James and the Giant Peach*, Dahl 76
E.3.49 *Automating the News*, Diakopoulos 77
E.3.50 *Drown*, Díaz 77
E.3.51 *The Brief Wondrous Life of Oscar Wao*, Díaz 78
E.3.52 *This Is How You Lose Her*, Díaz 78
E.3.53 *The White Album*, Didion 79
E.3.54 *Down and Out in the Magic Kingdom*, Doctorow 79
E.3.55 *The World's Wife*, Duffy 80
E.3.56 *A Visit from the Goon Squad*, Egan 80
E.3.57 *Invisible Man*, Ellison 81
E.3.58 *We Should All Be Mirandas*, Fairless & Garroni 81
E.3.59 *The Dude Abides*, Falsani 82
E.3.60 *The President's Vampire*, Farnsworth 82
E.3.61 *The Great Gatsby*, Fitzgerald 83
E.3.62 *Gone Girl*, Flynn 83
E.3.63 *British Destroyers*, Friedman 84
E.3.64 *Florals & Botanicals*, Fudurich et al. 84
E.3.65 *Good Omens*, Gaiman & Pratchett 85
E.3.66 *The Slippery Year*, Gideon 85
E.3.67 *Sweater Surgery*, Girard 86
E.3.68 *Blink*, Gladwell 86
E.3.69 *The Land Before Avocado*, Glover 87
E.3.70 *Dead Ringers*, Golden 87
E.3.71 *Ararat*, Golden 88
E.3.72 *Lord of the Flies*, Golding 88
E.3.73 *Wizard's First Rule*, Goodkind 89
E.3.74 *Rome and Jerusalem*, Goodman 89
E.3.75 *The Fault in Our Stars*, Green 90
E.3.76 *The Third Man*, Greene 90
E.3.77 *The Confessions of Max Tivoli*, Greer 91
E.3.78 *The Fugitive*, Grisham 91
E.3.79 *The Curious Incident of the Dog in the Night-Time*, Haddon 92
E.3.80 *Migrations to Solitude*, Halpern 92
E.3.81 *Uncommon Type*, Hanks 93
E.3.82 *Buzz*, Hanson 93
E.3.83 *Requiem for the Sun*, Haydon 94
E.3.84 *Catch-22*, Heller 94
E.3.85 *The Old Man and the Sea*, Hemingway 95
E.3.86 *Life on Air*, Hendy 95
E.3.87 *Great Hair Days*, Hersheson 96
E.3.88 *Graft*, Hill 96
E.3.89 *The Outsiders*, Hinton 97
E.3.90 *The Second Summoning*, Huff 97

E.3.91 *Selected Poems of Langston Hughes*, Hughes 98
E.3.92 *M. Butterfly*, Hwang 98
E.3.93 *Building and Operating a Realistic Model Railway*, Jackson 99
E.3.94 *All the Onions*, Jacobs 99
E.3.95 *Fifty Shades of Grey*, James 100
E.3.96 *The Stone Sky*, Jemisin 100
E.3.97 *Ulysses*, Joyce 101
E.3.98 *Sandman Slim*, Kadrey 101
E.3.99 *Ethnography after Antiquity*, Kaldellis 102
E.3.100 *Who Is Rich?*, Klam 102
E.3.101 *The Servants of Twilight*, Koontz 103
E.3.102 *Tai Chi for Depression*, Kuhn 103
E.3.103 *The Tide Was Always High*, Kun 104
E.3.104 *Girl in Translation*, Kwok 104
E.3.105 *A Wrinkle in Time*, L'Engle 105
E.3.106 *Call Me Brooklyn*, Lago 105
E.3.107 *Dead Wake*, Larson 106
E.3.108 *The Girl with the Dragon Tattoo*, Larsson 106
E.3.109 *The Daughter of Odren*, Le Guin 107
E.3.110 *The Chronicles of Narnia*, Lewis 107
E.3.111 *After I'm Gone*, Lippman 108
E.3.112 *Sunburn*, Lippman 108
E.3.113 *Marvel's Spider-Man: Hostile Takeover*, Liss 109
E.3.114 *Anastasia on Her Own*, Lowry 109
E.3.115 *Nixon in China*, MacMillan 110
E.3.116 *The Doomsday Prophecy*, Mariani 110
E.3.117 *A Game of Thrones*, Martin 111
E.3.118 *Rough-Hewn Land*, Meldahl 111
E.3.119 *Twilight*, Meyer 112
E.3.120 *The Duchess War*, Milan 112
E.3.121 *On Liberty*, Mill 113
E.3.122 *Coal Creek*, Miller 113
E.3.123 *Winnie the Pooh*, Milne 114
E.3.124 *Catching the Sky*, Moore & O'Brien 114
E.3.125 *The Heretic Queen*, Moran 115
E.3.126 *The Gondola Maker*, Morelli 115
E.3.127 *Songs in Ordinary Time*, Morris 116
E.3.128 *Beloved*, Morrison 116
E.3.129 *Norwegian Wood*, Murakami 117
E.3.130 *Eat More Plants*, Nielsen 117
E.3.131 *Polaris*, Northrop 118
E.3.132 *Pagans*, O'Donnell 118
E.3.133 *Windfall*, O'Sullivan 119
E.3.134 *The Memory Police*, Ogawa 119
E.3.135 *Winter Sisters*, Oliveira 120
E.3.136 *Nineteen Eighty Four*, Orwell 120
E.3.137 *Fight Club*, Palahniuk 121
E.3.138 *The Complete Joy of Homebrewing*, Papazian 121
E.3.139 *The Cult of Loving Kindness*, Park 122
E.3.140 *Along Came a Spider*, Patterson 122
E.3.141 *Payard Cookies*, Payard & McBride 123
E.3.142 *Essential Pepin Desserts*, Pépin 123
E.3.143 *Why New Orleans Matters*, Piazza 124
E.3.144 *Enlightenment Now*, Pinker 124

E.3.145 *Competitive Strategy*, Porter 125
E.3.146 *Night Watch*, Pratchett 125
E.3.147 *The Subtle Knife*, Pullman 126
E.3.148 *The Seductions of Darwin*, Rampley 126
E.3.149 *Kitchen Table Wisdom*, Remen 127
E.3.150 *Backroads Boss Lady*, Roberts & Witter 127
E.3.151 *Soft in the Head*, Roger 128
E.3.152 *The Making of a Mediterranean Emirate*, Rouighi 128
E.3.153 *Harry Potter and the Sorcerer's Stone*, Rowling 129
E.3.154 *Harry Potter and the Chamber of Secrets*, Rowling 129
E.3.155 *Harry Potter and the Goblet of Fire*, Rowling 130
E.3.156 *Harry Potter and the Deathly Hallows*, Rowling 130
E.3.157 *Born to Walk*, Rubinstein 131
E.3.158 *The Pretender*, Ruskin 131
E.3.159 *Toscanini*, Sachs 132
E.3.160 *Cosmos*, Sagan 132
E.3.161 *Middle India*, Sahni 133
E.3.162 *The Catcher in the Rye*, Salinger 133
E.3.163 *Lean In*, Sandberg 134
E.3.164 *The DevOps Adoption Playbook*, Sharma 134
E.3.165 *Frankenstein*, Shelley 135
E.3.166 *Sally Ride: America's First Woman in Space*, Sherr 135
E.3.167 *A Perfectly Good Family*, Shriver 136
E.3.168 *The Bedwetter*, Silverman 136
E.3.169 *On the Road with Bob Dylan*, Sloman 137
E.3.170 *The Night Children*, Smith 137
E.3.171 *White Teeth*, Smith 138
E.3.172 *No Visible Bruises*, Snyder 138
E.3.173 *The Grapes of Wrath*, Steinbeck 139
E.3.174 *Jesse James*, Stiles 139
E.3.175 *Zombie Halloween*, Stine 140
E.3.176 *Fear of Music*, Stubbs 140
E.3.177 *Pearl Harbor*, Swanson 141
E.3.178 *Chesapeake Requiem*, Swift 141
E.3.179 *The Goldfinch*, Tartt 142
E.3.180 *Unglued*, TerKeurst 142
E.3.181 *Embraced*, TerKeurst 143
E.3.182 *Birding with Yeats*, Thomson 143
E.3.183 *The Hobbit*, Tolkien 144
E.3.184 *The Fellowship of the Ring*, Tolkien 144
E.3.185 *Tree and Leaf*, Tolkien 145
E.3.186 *Noodles Every Day*, Trang 145
E.3.187 *Billionaire Democracy*, Tyler 146
E.3.188 *Portugal's Guerrilla Wars in Africa*, Venter 146
E.3.189 *Slaughterhouse-Five*, Vonnegut 147
E.3.190 *Animal Rights*, Waldau 147
E.3.191 *Men We Reaped*, Ward 148
E.3.192 *Charlotte's Web*, White 148
E.3.193 *A Return to Love*, Williamson 149
E.3.194 *Another Brooklyn*, Woodson 149
E.3.195 *Brown Girl Dreaming*, Woodson 150
E.3.196 *A Little Life*, Yanagihara 150
E.3.197 *The Art of Bonsai*, Yoshimura & Halford 151
E.3.198 *A People's History of the United States*, Zinn 151

E.3.199 *The Future of the Internet and How to Stop It*, Zittrain 152
E.3.200 *The Book Thief*, Zusak 152
E.4 Estimating how much of a book is memorized by a model 153
E.4.1 Quantifying extraction coverage 153
E.4.1 Extended extraction coverage results 153
F Measurement validity 160
F.1 Motivating our validity experiments 161
F.1.1 Setting τ_{min} and extending to models with unknown training data 161
F.1.2 Relating greedy-decoded and probabilistic discoverable extraction 162
F.1.3 The role of prefix length in our experiments 163
F.2 Setup 164
F.3 Results of varying prefix length on known training data 165
F.4 Results of negative control experiments 168
F.4.1 Results on PHI 4 169
F.4.2 Results on books from after training cutoffs 172
F.5 Additional notes on "the monkey at the typewriter" metaphor 174
F.6 Additional notes on extraction and membership inference 177
F.6.1 The relationship between membership inference and extraction 177
F.6.2 We know ground-truth membership for most of our experiments 177
F.6.3 Inferring membership for LLMs with undisclosed training data 178
F.6.4 We infer membership for sequences, not whole books 179
F.6.5 Absence of extraction does not imply anything about membership 180
F.6.6 Extractability and copyright 180
G Recovering books with one seed prompt 180
G.1 Recovery method 180
G.2 Partial recovery 184
G.3 Running recovery for books with low memorization signal 184
H Extended discussion of results 185
H.1 Copyright analysis 185
H.1.1 LLMs could be derivative works of the books they memorize 185
H.1.2 It may be practical to generate infringing copies in some cases 187
H.1.3 Class actions are ill-suited for copyright claims about memorization 188
H.2 Results on 200 books 188

A longer version of this paper is published on arXiv. (Cooper et al., 2025)

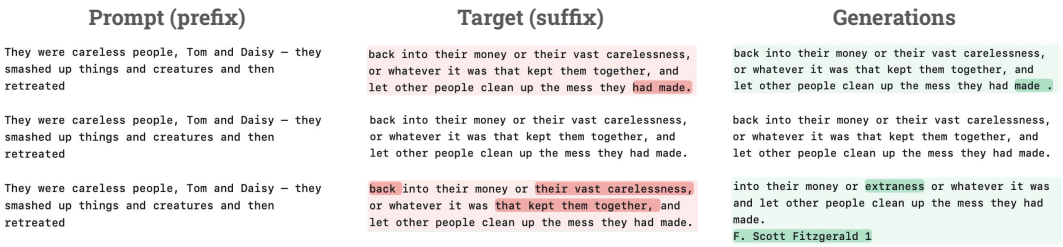


Figure 8: **Illustrating different possible completions of the "careless people" quote.** For a prefix of a sequence from *The Great Gatsby* (Fitzgerald, 1925), we show three 32-token completions when using the prefix as a prompt to LLAMA 1 30B (top-k, $T=1$ and $k=40$). Each row reflects a different interaction with the LLM: (**left**) the ground-truth prefix, used as the prompt (which is always the same); (**middle**) the target suffix, which is also always the same text; (**right**) the text of the generation in response to the prompt. For the target suffix, faint red highlighting indicates that LLAMA 1 30B did not reproduce the target suffix verbatim for the given interaction, with darker red highlighting indicating words from the target suffix that are missing in the generation. For generations, light green highlighting indicates that LLAMA 1 30B did not reproduce the target suffix verbatim for the given interaction, with darker green highlighted words reflecting text that the generation adds to the ground-truth target suffix. The middle row, where there is no highlighting, reflects verbatim extraction.

A Additional notes on memorization and probabilistic extraction

We provide additional notes on the background material in this paper, including the "careless people" quote in Section 2 (Appendix A.1), the key probabilistic extraction metric in this paper (formally called (n, p)-discoverable extraction, Appendix A.2), as well as our implementation for this metric (Appendix A.3). Please also refer to Section 2.

A.1 Additional notes on the "careless people" example

The headline figure we include in Figure 1 (Section 2) is meant to be illustrative—an instance of, not a general statement about, memorization and extraction. (In Figure 8, we also provide examples of outputs of prompting LLAMA 1 30B with the prefix for this example.) This example also plays a role in some of our validity experiments (Section 4 and Appendix F). The text is drawn from *The Great Gatsby* (Fitzgerald, 1925); it is one of the most famous quotes from the book. It is also a short, 57-token sequence: the prefix is 25 tokens (including the start-of-sequence token) and the suffix is 32 (with the LLAMA 1 tokenizer). We do not focus on such short suffixes in our analysis, for reasons that are discussed both in the main text and in our experimental details below. While extraction of verbatim 32-token suffixes is often accepted as valid evidence for memorization (Biderman et al., 2023), it is more common practice to use 50-token suffixes. (Regarding validity, also refer to Section 4 and Appendix F.) We compute the extraction probability with respect to top-k decoding ($T = 1$, $k = 40$). We do this by transforming the model's output logits according to the sampling parameters of this scheme. (See more about this below in Appendix A.3.)

It is also the case that very famous quotes like this are duplicated in many locations in training-data sources (e.g., blogs, essays, personal web pages, etc.) We are not claiming (here, or in general) that this sequence can be extracted *because* a given whole book (here, *The Great Gatsby*) was included in `Books3`; it could be extracted (or more extractable) because it is duplicated in many places in the training data. Of course, one of those copies is in `Books3` for LLMs where the entirety of `Books3` is included in the training data. We discuss this further in Appendix F.

A.2 Metrics

The metric that we use for quantifying memorization comes from Hayes et al. (2025b). The authors provide a different framing for the metric than we do. Instead of directly comparing p_z values (for the same sequence across LLMs, across multiple sequences for the same LLM,

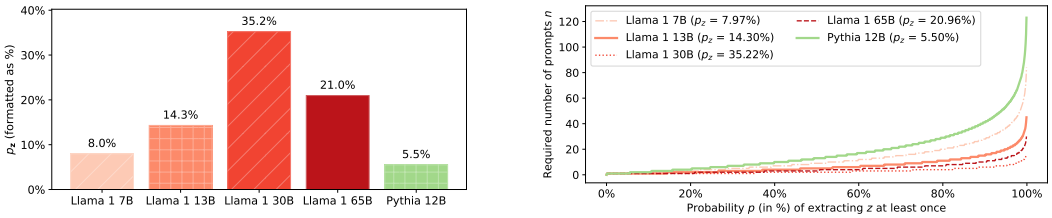


Figure 9: **Plotting p_z for the "careless people" quote from *The Great Gatsby*.** (**left**) p_z (1) for different models. (**right**) Translating p_z into how many prompts n it would take to extract the sequence z with at least probability p (3). See Section 2.

etc.), Hayes et al. (2025b) specify a threat model with an adversary that can query the model an arbitrary number of times with the same prompt—an adversary that has n independent attempts to extract a sequence z (its suffix) with at least probability p. This is where their metric derives its name, **(n, p)-discoverable extraction**:

Definition A.1 (**(n, p)-discoverable extraction**, from Hayes et al. (2025b))**.** Given a training sequence z that is split into an a-length prefix $z_{1:a}$ and a j-length suffix $z_{a+1:a+j}$, z is (n, p)-*discoverably extractable* if

$$\Pr\left(\cup_{w\in[n]}\,(g_\phi \circ f_\theta)^j_w(z_{1:a}) = z_{1:a+j}\right) \geq p,$$

where $(g_\phi \circ f_\theta)^j_w(z_{1:a})$ represents the w-th (of n) independent execution of the autoregressive process of generating a distribution over the token vocabulary, sampling a token from this distribution, and adding the token to the sequence $j > 0$ times, starting from the same initial sequence $z_{1:a}$.

On its own, this type of method is well-suited to **completion** models that continue the prompt text, rather than instruction-tuned **chatbots** that reply in a conversational style. Similar to Hayes et al. (2025b) (and other work on discoverable extraction), we therefore only study completion models and, for brevity, refer to these models by name, version, and size. Also note that this definition can be (at least in principle) easily adapted to non-verbatim extraction by testing if a generation is within ϵ distance from the target suffix for a given distance metric. Despite these differences in framing, in practice for verbatim extraction, computing this metric just involves calculating sequence probabilities p_z. For sequence z, this is the suffix probability (for the given prefix) with respect to an LLM's base distribution θ and decoding scheme ϕ (e.g., greedy, top-k) that determines which tokens are considered at each step and how their logits are renormalized into conditional probabilities (Appendix A.3). That is,

$$\begin{aligned} p_z := p_{z,\theta,\phi} \triangleq \Pr_{\theta,\phi}(z_{a+1:a+j} \mid z_{1:a}) &= \prod_{t=a+1}^{a+j} \Pr_{\theta,\phi}(z_t \mid z_{1:t-1}) \\ &= \exp\left(\sum_{t=a+1}^{a+j} \log \Pr_{\theta,\phi}(z_t \mid z_{1:t-1})\right). \end{aligned} \tag{3}$$

We always note the particular LLM θ and decoding scheme ϕ, and so, for brevity, will refer to this quantity as simply p_z. In practice, we compute p_z as the exp of the sum of conditional log probabilities of the tokens in the sequence, as this is more numerically stable than multiplying together the conditional probabilities of the tokens in the sequence.

In Figure 9 (left), we show examples of p_z for the "careless people" quote (split into prefix and suffix as in Figure 1) for different models. (This is Figure 1, right, in Section 2, with PYTHIA 12B added.) Once p_z is, one can compute n for a given p (for $p_z \in (0, 1)$) with

$$1 - (1 - p_z)^n \geq p \implies n \geq \frac{\log(1-p)}{\log(1-p_z)}. \tag{4}$$

Since p_z is the probability of generating the suffix with 1 prompt to the model, the probability of *not* generating it in 1 prompt is $(1 - p_z)$. The probability of *not* generating it

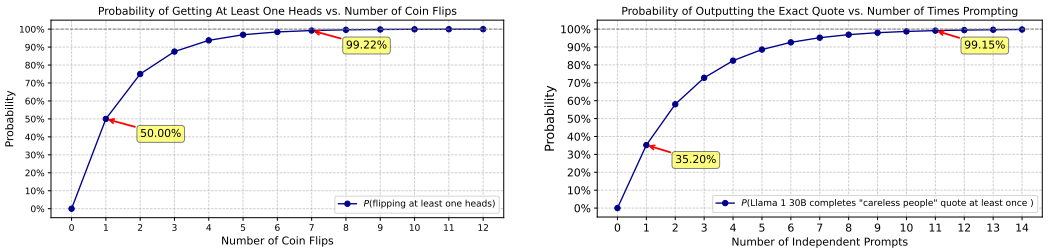


Figure 10: **Probability of an event occurring *at least once* changes as a function of n independent trials.** Following the intuition of flipping a fair coin (where heads has $p_z = 0.5 = 50\%$), we show how the probability of flipping heads at least once changes with more flips (**left**). We show how the probability p of generating the verbatim suffix of the "careless people" quote at least once for LLAMA 1 30B ($p_z \approx 35.2\%$) changes as a function of the number of independent prompts to the model with the prefix (**right**). Both figures use $1 - (1 - p_z)^n$ (Equation 4), plugging in the respective p_z (for the coin flip and for generating the "careless people" suffix given the prefix) and different values of n to compute p.

in n independent prompts to the model (with the same prompt) is $(1 - p_z)^n$, and so the probability of generating it in n prompts is 1 minus this probability, i.e., $1 - (1 - p_z)^n$.

As an intuition for this presentation of the metric, think about a fair coin flip coming up heads. Here, $p_z = 0.5$ (the probability of heads in one flip). With two actual flips ($n = 2$), the probability of flipping heads *at least once* is $1 - (1 - 0.5)^2 = 0.75 = p$; when $n = 10$, $p = 0.9990$. We visualize this in Figure 10 (left) with coin flips, and show how it translates to suffix generation probabilities for LLAMA 1 30B and the "careless people" quote, for which $p_z = 35.2\%$ (right). As is clear Equation 4, for a given p_z, one can pick a probability threshold p and get the corresponding number of prompts n (or vice versa). Figure 9 (right) visualizes how n changes for different settings of p for the "careless people" quote.

In this paper, we re-frame (n, p)-discoverable extraction to emphasize the quantity p_z (Equation 3) that we actually compute: the probability of extracting a sequence $\boldsymbol{z}$ (for the given LLM, hyperparameter-configured decoding scheme, and suffix start index location within sequence $\boldsymbol{z}$). This is also useful for plotting distributions over p_z for a given book—to see how probabilities vary for sequences across a book, or for a given sequence across different models. We provide such comparisons in Section 5.

In practice, we can compute p_z for verbatim extraction *without* prompting n times. We compute sequence probabilities directly from the logits we obtain from running whole sequences through the LLM, rather than generating multiple suffixes for each prefix. (See Appendix A.3.) These two approaches are equivalent for verbatim extraction: the computed suffix probabilities correspond to the statistically expected frequencies of verbatim extracted outputs we would observe if we were to generate a large number of continuations for the same prompt. Hayes et al. (2025b), the authors of this extraction approach, confirm this with ample experimental evidence. We do, as well, though omit these results for brevity because they follow directly from what p_z quantifies. This is also why we also limit ourselves to verbatim extraction in this paper; the efficiencies of operating directly on logits with one forward pass only apply to this setting. (See Section 2 and Appendix A.3 for more details.) As a result, our work here (in several respects) only scratches the surface: there are various different ways to instantiate this extraction methodology in practice, we only explore a limited set of open-weight models, and we only test 200 of the nearly 200,000 books in `Books3`.

We discuss how we determine extraction success for probabilistic extraction in Appendix D. In the main text, we discuss determining extraction success in Section 2.

A.3 Computing sequence probabilities in one forward pass

Our implementation differs from Hayes et al. (2025b), which we confirm in discussion with the authors. The original paper generated the suffix a token at a time and summed up the per-generated-token conditional log probabilities. Here, we observe that this is not necessary.

We can get the logits for each token in the sequence with just one forward pass through the model, and effectively only one line of inference code. This also makes computing this metric cheaper than traditional greedy-decoded discoverable extraction (Carlini et al., 2021), in terms of wall clock time; one large matrix-matrix multiplication is significantly cheaper than generation with KV-caching. Note that this is the *entire* interaction with the model in our main experiments. We are not making *any* changes to the model in our extraction measurements.

We then implement our sequence probability computations (i.e., compute p_z) as a post-processing operation on the logits. For any given decoding scheme, we can transform the logits—for example, change the shape according to temperature, truncate them to the top-k—before normalizing them into the log probability distribution. This computation is written out on the right side of Equation 3. The key point here is that, for a given sequence that we are evaluating, all we do (starting at the first index in the suffix) is sum up the conditional log probabilities of the *actual* next tokens in the sequence, one at a time. This is often called **teacher forcing**.

We work with the conditional log probabilities (adding them together), rather than the conditional probabilities (and multiplying them together), as this is more numerically stable to compute. For example, consider the sequence `"the orange cat"` (where, for simplicity, we will assume each word is a single token). When processing `"the"`, the logits reflect the unnormalized scores of the next token. We find the logit value that corresponds to `_orange`, compute log softmax to get the log probability, and add the associated conditional log probability. Then we continue on (doing the same for `_orange`, where we get the conditional log probability for the next token being `_cat`). We get the conditional log probability of the whole suffix by adding together the conditional, per-token log probabilities of the actual suffix. From the suffix log probability, we easily obtain the probability p_z.

Any given continuation should be low probability. That is, for a suffix length of 50 tokens, this computation is effectively multiplying together 50 per-token conditional probabilities (by adding together conditional log probabilities). For a continuation where each token has really high probability, this can still mean the overall continuation has relatively low probability. Consider that each token in a suffix has conditional probability 0.9. This is *really* high; it means that a single token in the logit vector (at each position in the suffix) has 90% of the corresponding probability mass; for a token vocabulary of size $32,000$, that means the remaining $31,999$ tokens all share the remaining 10%. For this suffix, the probability of generating it is $0.9^{50} \approx 0.005$—i.e., a 0.5% chance of generation. This is a relatively small number in the scheme of things; but it is also a really large number in general, if we consider what this means for the underlying token probabilities. In some cases in this paper, we observe sequences that have ***over*** 90% ***probability***. Such unusually high probabilities associated with verbatim suffixes in the training data are indicative of memorization (where it is reasonable to consider 0.5%, or even smaller, to be unusually high).

Also note that top-k sampling truncates the logit vector to *only* the top-k tokens (e.g., for a vocabulary of 32,000 tokens, there are 32,000 logits, but top-k with k=40 decoding will only consider the 40 highest-valued logits). This means that, if a given suffix has a token at a position that is *not* in the top-k of the logit vector, then we will not be able to complete the probability computation. In this case, the suffix has $p_z = 0$; it is not extractable with *any* probability.

A refactored version of our experimental code, with additional optimizations for increased efficiency, can be found here.

B Background on memorization, extraction risk, and U.S. copyright

The dozens of pending copyright lawsuits concerning LLMs (Chat GPT Is Eating the World) generally present three interrelated issues: (1) whether training an LLM on copyrighted material is permitted **fair use** (i.e., limited use of the copyrighted material can, under certain circumstances, be used without permission from the copyright owner); (2) whether the model itself is a **copy** or **derivative work** (technical terms in copyright law) of the

works on which it is trained (Section 7 & Appendix H.1); and (3) whether the model reproduces copyrighted material in its outputs. Some suits present only one issue, while others present all three questions. Some suits are based on content owned by a single company, while others are class action lawsuits purporting to represent all book authors.[3] These lawsuits have primarily focused on (1) training and on (3) model outputs, and the two decisions so far in the U.S. have held that training an LLM can be fair use, albeit with significant limitations (Kadrey Judgment; Bartz Judgment). They also raise issues about the acquisition, storage, or use of **shadow libraries** of copyrighted material as training data, such as `Books3` (Reisner, 2023; Lee et al., 2023b).

Our work does not speak to the first issue; training and fair use have been discussed extensively elsewhere (Lemley & Casey, 2021). Rather, our findings affect the other two issues: we show that copies are sometimes made in the model itself (i.e., memorized) and those copies can sometimes be generated as outputs (i.e., extracted). Because many LLMs are known to have been trained on `Books3`, we use books from this dataset to study memorization and extraction of copyrighted material (Sections 3–5; Appendix E).

Memorization and copies inside the model. If an LLM memorizes all or a substantial portion of a copyrighted work (near-)verbatim, the LLM itself may be or contain an infringing copy or derivative work (Lee et al., 2023b; Cooper & Grimmelmann, 2024). This is because memorized training data are encoded inside the model (Sections 1 & 2). Others have argued that encoding the work in the form of model weights satisfies the technical definition of "copy" in the U.S. Copyright Act (Cooper & Grimmelmann, 2024). In a technical sense, that copy may be of interest to copyright even if there is no practical way to access or extract it. Like the copies of copyrighted works used as training data during training, that copy in model might be fair use, but the analysis of fair use would look somewhat different than *internal* use of copyrighted works in training processes. This is particularly true for open-weight models like DEEPSEEK and LLAMA models, which are not merely used internally by their developers, but are themselves shared *externally* with others (Section 7 & Appendix H.1).

Extracting copies in model outputs. Because some memorized works in the training data are extractable, LLMs can sometimes output copies of those works to users (Cooper et al., 2023; Lee et al., 2023b; Henderson et al., 2023; Lee et al., 2024; Cyphert, 2024; Sag, 2024b; Charlesworth, 2025). Works with higher extraction probabilities exhibit greater extraction risk; it takes fewer attempts to produce an output that is a verbatim copy of such a work (Section 2 & Appendix A.2). That output at generation time will be judged separately from the model itself, and is less likely to be a fair use (Henderson et al., 2023; Lee et al., 2023b).

Extraction is inherently probabilistic under non-deterministic decoding schemes (Section 2), so the relevant question is not only *whether* a copy could ever appear for some prompt, but also *how often* that copy appears in order to be "meaningful" (Cooper & Grimmelmann, 2024). When we measure suffix probabilities p_z that are very close to 0%, we do not count those sequences as extracted; one could perhaps understand such generations as instances of "a monkey at the typewriter" (Borel, 1913; The Simpsons: Last Exit to Springfield (Fox television broadcast Mar. 11, 1993)) (Section 4). Very large p_z (e.g., $>90\%$) indicates near-certain reproduction; this is clearly extraction (and therefore evidence of memorization). However, most cases are not this clear, and it is not immediately obvious where to draw a line in the enormous range between these extremes. To avoid false positives, we use a conservative threshold $\tau_{\text{min}} = 0.1\%$ for *scientific* evidence of extraction (Sections 2 & 4; Appendices A.2 & F). But for *copyright law and policy*, the practicability of extraction may require a different bar: if reproducing a work just once would require thousands of prompts, the real-world risk may be relatively limited, even if nonzero (Section 7).[4]

[3] There are also other copyright issues (Lee et al., 2023b; Samuelson, 2023; Sag, 2024a; Goodyear, 2025; Sobel, 2024; Cooper et al., 2024), for example, whether novel outputs of generative-AI systems are themselves copyrightable (Lee et al., 2023b; United States Copyright Office).

[4] Relatively low extraction probabilities (e.g., $p_z = 0.1\%$) do not always mean that it would take a large number of attempts to extract the suffix. In some cases, other decoding schemes like beam search may make extraction of such suffixes deterministic.

Altogether, setting sensible copyright policy requires a specific understanding of both of these sets of issues. However, ML research does not tend to investigate these types of work-specific questions. We begin to address this gap by quantifying memorization of individual books.

C Testing our measurement pipeline

To confirm the accuracy of our implementation of probabilistic discoverable extraction, as well as to benchmark its efficiency in comparison to the implementation in Hayes et al. (2025b), we re-run the experiments from Hayes et al. (2025b) for small PYTHIA models on the `Enron` dataset. To ensure consistency, we discussed the original run times and sample of `Enron` with the authors of Hayes et al. (2025b). We provide some brief details about these experiments to show due diligence for our implementation in our particular setting.

We ran these experiments with both `float16` and `float32`, and report results for `float32` to align with Hayes et al. (2025b). (However, we note no significant differences between the two.) In Table 1, we show the runtime for our experiments on the 10,000, 100-token sequences Hayes et al. (2025b) drew from `Enron`. In Table 2, we report the same metrics as Hayes et al. (2025b): the greedy-decoded discoverable extraction rate and (as a point of comparison) the maximum probabilistic rate. We run these experiments on the same 4 A100s as all of our other probabilistic extraction experiments.

Table 1: **Runtime for PYTHIA on `Enron`.** We run 2 PYTHIA models on the same 10,000 100-token sequences that Hayes et al. (2025b) use from the `Enron` dataset.

	Time (mm:ss)	Parallelism	Batch size
PYTHIA 1B	02:30.42	Data (four GPUs)	250
PYTHIA 2.8B	03:36.35	Data (four GPUs)	250

Table 2: **Comparing our results to Hayes et al. (2025b).** We give the greedy discoverable extraction and maximum probabilistic rate for PYTHIA models on the 10,000-sequence subset from `Enron`. We successfully replicate the exact results in Hayes et al. (2025b), which did not make use of a beginning of sequence (`BOS`) token. We compare to the rates where the `BOS` token is included at the start of the sequence. Note that the probabilistic rate increases slightly with the presence of the `BOS` token.

PYTHIA 1B on Enron subset	**Hayes et al. (2025b) (No `BOS` token)**	**With `BOS` token**
Greedy extraction rate	0.76%	0.74%
Max. probabilistic rate	5.27%	5.52%
PYTHIA 2.8B on Enron subset	**Hayes et al. (2025b) (No `BOS` token)**	**With `BOS` token**
Greedy extraction rate	1.3 %	1.82%
Max. probabilistic rate	9.04%	9.47%

We successfully replicate the results of these experiments in Hayes et al. (2025b). However, in the process of doing so, we identified a small bug in their original implementation. The PYTHIA tokenizer is a GPT-2-style tokenizer, and does *not* add a beginning of sequence (`BOS`) token by default to the sequences it tokenizes. However, for our setting, it is best practice to include this token at the start of a sequence, as we submit these sequences as inputs for inference to LLMs (that are trained to expect this token at the beginning of the sequence). When we manually prepend this token and re-run the experiments from Hayes et al. (2025b), we generally observe elevated rates compared to Hayes et al. (2025b). See Table 2. We similarly prepend the `BOS` token in our `Books3` experiments using PYTHIA models.

D Experiments on extraction rates

Similar to other extraction papers, we also compute average extraction rates. These serve as a reference point for our discussion concerning how average rates can conceal interesting patterns of work-specific memorization. In the main text, we refer briefly to these experiments in Section 3.2.

Computing average extraction rates. Let $\mathbb{Z}$ be a set of sequences z drawn from some (part of a) training dataset. Each z is split into a prefix $z_{1:a}$ and target suffix $z_{a+1:a+j}$ (e.g., $a{=}j{=}50$ for 100-token sequences).

All extraction rates take the form

$$\text{extraction\_rate}(\mathbb{Z}; s) \;=\; \frac{1}{|\mathbb{Z}|} \sum_{z \in \mathbb{Z}} \mathbf{1}[s(z)], \tag{5}$$

where $s(z)$ is a success predicate defined according to the chosen criterion.

Fix a model θ and a decoding scheme ϕ (e.g., base distribution $T{=}1$, or top-k with $k{=}40$), which induces a conditional next-token distribution that can be used to compute suffix probabilities $p_z \in [0,1]$. For a sequence z with prefix $z_{1:a}$ and suffix $z_{a+1:a+j}$, define the suffix probability under θ and ϕ as

$$p_z \coloneqq p_{z,\theta,\phi} \;=\; \prod_{t=a+1}^{a+j} \Pr_{\theta,\phi}(z_t \mid z_{1:t-1}).$$

This is the probability that the model generates the exact target suffix token by token under decoding scheme ϕ; it is the same definition as Equation 3 that we use in Section 2 and Appendix A.

For traditional, greedy-decoded discoverable extraction (Carlini et al. (2023); Lee et al. (2022); Section 2), $\phi{=}\,\text{greedy}$ (i.e., $k{=}1$) and the success predicate can be written simply as explicit string equality:

$$s_{\theta,\text{greedy}}(z) \;=\; \mathbf{1}\big[\,\text{greedy}(\theta; z_{1:a}) = z_{a+1:a+j}\,\big]. \tag{6}$$

More generally, for any LLM θ and decoding scheme ϕ, the success predicate just thresholds the suffix probability. Given a minimum threshold $\tau_{\text{min}} \in (0,1]$,

$$s_{\theta,\phi}(z; \tau_{\text{min}}) \;=\; \mathbf{1}\big[\,p_{z,\theta,\phi} \geq \tau_{\text{min}}\,\big]. \tag{7}$$

Again, note that greedy decoding is a special case: each top-1 token has probability 1 under $\phi = \text{greedy}$, so $p_{z,\theta,\phi} = 1$ and success reduces to equality with the suffix. In the main paper, we omit θ and ϕ in our notation. We always mention which specific model we are evaluating, and we are always considering ϕ to be top-k decoding with $T{=}1$ and $k{=}40$. (See Appendix E.2 for more details on this choice.) When using non-deterministic decoding schemes and suffix probabilities to quantify extraction, we very conservatively set $\tau_{\text{min}}{=}0.1\%$ (Section 4 & Appendix F).

D.1 Setup

Data. We perform our average extraction rate experiments with the `Books3` dataset. We obtained this dataset from a previous (2022) download of The Pile (Gao et al., 2020), which is stored on a university cluster and which we use for research purposes on language modeling. The status of `Books3` as a research artifact (as well as a training corpus for LLMs) remains unresolved. (See Appendix B.) Nevertheless, this dataset remains widely available in the research community; for example, it can be found in the HuggingFace-hosted version of The Pile (EleutherAI/the_pile_deduplicated), as well as in various other

data repositories (e.g., SaylorTwift/the_pile_books3_minus_gutenberg; amongglue/books3-subset-raw; CANBERT/pile_books3_text). It also features as part of a benchmark task in the popular HELM evaluation suite. For this set of experiments, we draw random samples from Books3, performing 3 separate runs each on 40,000 sequences. We discuss how we sample these sets of sequences in Appendix D.2.

Models. From 13 different models, we try to extract 40,000 sequences from Books3 (3 times each). We test LLAMA 1 7B, LLAMA 1 13B, LLAMA 1 65B (Touvron et al., 2023a), LLAMA 2 7B, LLAMA 2 13B, LLAMA 2 70B (Touvron et al., 2023b), LLAMA 3.1 8B, LLAMA 3.1 70B (Grattafiori et al., 2024), PYTHIA 6.9B, PYTHIA 12B (Biderman et al., 2023), DEEPSEEK V1 7B (DeepSeek-AI et al., 2024), GEMMA 2 9B (Team et al., 2024), and PHI 4 (Abdin et al., 2024) (a 14B model). We group these models into three size groups: small, medium and large. We use `float16` for all models except for LLAMA 3 and LLAMA 3.1 models (and QWEN 2.5 models in sliding-window experiments). These models were explicitly trained to work with `bfloat16`.

We know for certain that PYTHIA and LLAMA models were trained on Books3, and can conclude from our validity experiments (Section 4 and Appendix F) that DEEPSEEK V1, QWEN 2.5 and GEMMA 2 models were trained on (at least parts of) some books that are also contained in Books3. PHI 4 was deliberately not trained on copyrighted (whole) books, but we find that we can still extract some snippets from copyrighted books. Unspecified public domain books were included in its training data, as well.

Metrics. We compute extraction rates using Equation 5 for both traditional, greedy-decoded discoverable extraction (using the success condition in Equation 6) and probabilistic extraction with top-k ($T = 1$, $k = 40$) decoding (using the success condition in Equation 7) and $\tau_{\text{min}} = 0.1\%$.

Compute resources. We run all of our average extraction rate experiments in a Slurm cluster environment, using the same node with 4 A100 GPUs.

D.2 Experimental procedure

Sampling books from Books3. To collect 40,000 sequences for which we compute an extraction curve for a given model, we use the following procedure. We sample 4,000 books without replacement from Books3. Then, from each of these books, we sample 10 non-overlapping 100-token sequences (50-token prefixes + 50-token suffixes). We make sure that the start of the prefix text is a space (to prevent unusual tokenization). We run these 40,000 sequences through each of the 13 models to compute extraction metrics. We follow this procedure 3 separate times (i.e., for a total of 120,000 sequences across 12,000 different books) to get a sense of the variance across extraction metrics for different (empirical) population estimates. Each run of 40,000 sequences took between approximately 8 minutes and 45 minutes on 4 A100s. We omit detailed timing results for brevity.

Probabilistic extraction configuration. We report greedy-decoded discoverable extraction and probabilistic discoverable extraction with top-k ($T = 1$, $k = 40$) and $\tau_{\text{min}} = 0.1\%$. The maximum probabilistic rate (also reported in Appendix C) should be understood as a very loose upper bound on extraction of memorized training data; it counts all sequences with non-zero p_z (i.e., it also includes sequences for which $p_z < \tau_{\text{min}} = 0.1\%$). For sequences with these low $p_z < \tau_{\text{min}}$, it may also capture generalization. We include these numbers as a reference, not to make extraction claims.

D.3 High-level takeaways

We defer discussion of validity to Appendix F, where we talk about how to draw conclusions for different models—some for which we know with certainty that Book3 is in their training data, and others for which we can conclude that at least some content that overlaps with Books3 was in their training data. PHI 4 reflects a separate case: it as trained predominantly

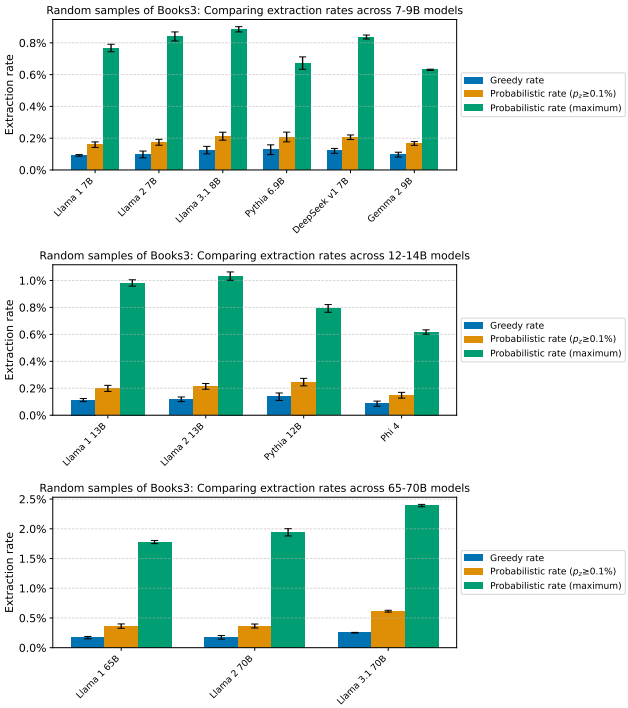


Figure 11: **Extended extraction rate results.** Comparing greedy discoverable and probabilistic extraction rates (for a our conservative setting, $p_z \geq \tau_{\text{min}} = 0.1\%$). As a reference point, we also include the maximum possible rate of generating 50-token sequences that match our suffixes drawn from `Books3` for small (7–9B), medium (12–14B), and large models (65–70B). (This is akin to looking at all $p_z > 0$.) Probabilistic rates are computed with top-k sampling using $T = 1$ and $k = 40$. The maximum probabilistic rate should be understood as a very loose upper bound on extraction of memorized training data; it also captures generalization. *Average* extraction rates (i.e., $\tau_{\text{min}} = 0.1\%$) on randomly sampled `Books3` sequences are relatively low for all models. Aligning with prior work, we observe that larger models memorize more. For our setting, this is true for both within and across model families. PHI 4, a 14B model, is an exception: its extraction rates are on par with smaller models. Error bars here indicate variance across the 3 samples of 40,000 sequences that we run.

on synthetic data. Based on the PHI 4 technical report (Abdin et al., 2024), PHI 4 was deliberately not trained on copyrighted (whole) books, but was trained on unspecified public domain material. As a result, `Books3` (the specific corpus) was almost certainly not included in its training data. Nevertheless, we are sometimes able to extract data that is contained in `Books3`, both from in-copyright and public domain books. Sequences from in-copyright texts likely found their way into the training data from web-scraped sources that quote these books. We discuss this further in Section 4, Appendix F, and Appendix H.

Figure 11 shows average extraction rates grouped roughly by model size. As we expected (and discuss in Appendix B), *average* extraction rates are relatively low. Using the conservative threshold of $p_z \geq \tau_{\text{min}} = 0.1\%$, the lowest rate we observe is for PHI 4 ($< 0.2\%$), and the highest is for LLAMA 3.1 70B ($\sim 0.6\%$). Consistent with prior work, our results show that larger models tend to memorize more than smaller ones (Hayes et al., 2025b; Carlini et al., 2023; Nasr et al., 2023; 2025). This holds within the same model family. PHI 4 is an exception; despite being a 14B model, its extraction rates on `Books3` sequences are comparable to the smaller (7–9B) models we evaluate. Small LLAMA 3 models memorize on-par with medium LLAMA 1 and 2 models. In general, later generations of LLAMA model families tend to memorize more `Books3`-contained text. For just one example, LLAMA 3.1 70B memorizes more on average than LLAMA 2 70B, which memorizes more

on average than LLAMA 1 65B. This pattern does not always hold for specific books. We defer additional, related observations about LLAMA 3.1 70B to Appendix H.

Altogether, it is important to keep in mind that any given sequence that we deem extracted is, by definition, a member of the training data; however, for LLMs with unknown training data, it is not necessarily a member of the training data *because* it came from the Books3 corpus. The sequence may be part of another source (or multiple other sources) in the training data. For PYTHIA and LLAMA models, the Books3 copy of a sequence may just be one of several duplicate copies in the whole training dataset. The same can be said for models like GEMMA 2, where we are not sure if the Books3 corpus was included in the training dataset.) Nevertheless, it is clear from our results that sequences that are contained in Books3—but perhaps originate from other sources—are included in the training data (Section 4 & Appendix F).

E Extended results for sliding-window experiments

We discuss details for our main experiments that use a sliding window across a book to surface memorization "hot-spots" at particular book locations. We discuss the overall setup (Appendix E.1), our sliding-window procedure (Appendix E.2), detailed results for 200 books and 14 LLMs (Appendix E.3), and our extraction coverage metric (Appendix E.4). We describe other results—beyond our main experiments with 50-token prefixes and 50-token suffixes—in different appendices. We defer discussion of sliding-window validity experiments on non-training data and different prompt prefix lengths to Appendix F. Additional experiments concerning extraction rates and our long-form extraction procedure are discussed in Appendix D and Appendix G, respectively.

E.1 Setup

Data and book selection. We draw our prompts (prefixes) from a selection of books in Books3; we verify extraction against the corresponding suffixes drawn from Books3. Books3 is a common pretraining dataset that is known to have been used to train several open-weight LLMs (Appendix D.1, Reisner (2023); Lee et al. (2023b). (See models, below.) We list the 200 selected books in Table 3.

We began our project with *The Great Gatsby* (Fitzgerald, 1925). We then continued our selection process with the books listed with the associated plaintiffs in the (amended) class action complaint of *Kadrey et al. v. Meta, Inc.* (Kadrey v. Meta, pp. 4-5). This is how we sourced the first 13 in-copyright books from Books3. We expanded to additional books by these plaintiffs, and also added in some generally popular books that we chose among the project team (e.g., Rowling (1997); Tolkien (1937); Camus (1942); Heller (1961)) and some (less publicly popular) academic books based on our personal preferences (e.g., Zittrain (2008); Barolini (2006)). (This team, after all, is composed of academics.) We deliberately included selections from the public domain (Carroll (1871); Fitzgerald (1925); Joyce (1922)), and an example of a book that was published under a permissive CC license (Doctorow (2003)).We added in Carroll (1871) because of concurrent work that was published on arXiv just as we posted our own (Ma et al., 2025). To round out the list, we also selected books at random from the manifest file that accompanies the Books3 archive (e.g., Jacobs (1993), Rouighi (2011), Cogburn & Silcox (2012)). Once we observed that certain very popular books were highly memorized by LLAMA 3.1 70B, we added additional bestsellers to our set (e.g., Martin (1996), Brown (2003), Meyer (2005), Orwell (1949), Sandberg (2013)). In earlier versions of this work, we reported on 36 and then 50 books in total.

In the current version, we manually selected additional books (non-randomly) and randomly sampled additional books, such that the final split contains 100 chosen books and 100 that were randomly sampled. Given that there are nearly 200,000 books in Books3, our experiments cover roughly 0.1% of the corpus. This is not nearly a large enough sample to make general claims about overall books extraction for the setup we study. This is why we do not completely randomly select books; we pick an assortment of manually and randomly selected books, intended to capture variation in the corpus.

Books metadata curation. In earlier versions of this project, which involved a smaller number of books, we manually gathered books metadata and manually created bibtex entries. When scaling up to 200 books, we attempted to automate as much as possible. For each book we evaluate in `Books3`, we used GPT-5 to extract bibliographic metadata, including the title and author(s), as well as the ISBN(s) and publisher(s) corresponding to the specific edition. To do so, we provided GPT-5 with the first and last several thousand words of each text, reflecting our observation that the desired metadata tended to concentrate in the front and back matter. Using these fields, we then queried GEMINI 2.5 FLASH with Google Search to determine each book's current copyright status and initial year of publication. Metadata quality was assessed through manually spot-checking a random 10% subset of the books. We found some errors in how authors' names were parsed, and so manually validated the author field for all 200 books. Refer to Table 3 for the gathered books metadata.

Models. Altogether, we run this set of experiments on 14 completion-style (i.e., non-chatbot) models: PYTHIA 12B (Biderman et al., 2023), LLAMA 1 13B, LLAMA 1 65B (Touvron et al., 2023a), LLAMA 2 7B, LLAMA 2 13B, LLAMA 2 70B (Touvron et al., 2023b), LLAMA 3 8B, LLAMA 3 70B, LLAMA 3.1 8B, LLAMA 3.1 70B (Grattafiori et al., 2024), DEEPSEEK V1 67B (DeepSeek-AI et al., 2024), QWEN 2.5 72B (Qwen et al., 2025), GEMMA 2 27B (Team et al., 2024), and PHI 4 (a 14B model) (Abdin et al., 2024). In earlier versions of this work, we also ran experiments on 3 other (smaller) models, which we omit for brevity: PYTHIA 6.7B, LLAMA 1 7B, and DEEPSEEK V1 7B.

We load all models with `float16`—except for LLAMA 3, LLAMA 3.1, and QWEN 2.5 models. These models were explicitly trained to work with `bfloat16`. In the first version of this work, we reported results for LLAMA 3.1 that used `float16`. Slight differences in results in this version are attributed to the change to `bfloat16`. (We have otherwise confirmed our results are deterministic, with respect to the hardware that we use throughout.) Regardless, the overall results are similar, and the conclusions remain the same.

We picked these models because they each fit (with respect to 16-bit format) on 4 A100s, and because they fall into three different categories that we wanted to investigate: (1) models that we know with certainty were trained on `Books3`, (2) models where we have reason to believe were trained on copyrighted books (whether drawn directly from `Books3` or from other sources), and (3) a model that we trust was not trained on whole copyrighted books.

For the first category, when we generate verbatim completions for these models on `Books3` data, we can very reasonably conclude that we have extracted data that these models have memorized during training. We know that these data were members of the training data; extraction success (and thus memorization) is a standard conclusion in this setting (e.g., Biderman et al., 2023; Hayes et al., 2025b; Prashanth et al., 2024; Carlini et al., 2023). For the second category, we aim to do valid membership inference of copyrighted data. The third category is included for all books, but is intended to help confirm validity; as a baseline, we intentionally include an LLM for which we do not expect to be able to extract much (if any) copyrighted books data.

1. **Category 1: PYTHIA 12B, LLAMA 1 13B, LLAMA 1 65B, LLAMA 2 7B, LLAMA 2 13B, LLAMA 2 70B, LLAMA 3 8B, LLAMA 3 8B, LLAMA 3 70B, and LLAMA 3.1 70B.** We know that PYTHIA and LLAMA models were trained on `Books3`. PYTHIA was trained on the Pile (Gao et al., 2020) (which contains `Books3`), based on EleutherAI's extensive documentation. We include only the largest (12-billion parameter) PYTHIA model.

   LLAMA 1 and LLAMA 2 are known with absolute certainty to have been trained on `Books3`. The LLAMA 1 report explicitly says so. It came out in discovery during Kadrey v. Meta Platforms, Inc. that LLAMA 2 models were all also definitively trained on `Books3`.

   Both in documents in discovery and Meta's otherwise public documentation, it is repeatedly noted that the development of LLAMA 3 continued to use the same data practices as prior versions. For minor-versioned LLAMA 3 models (like LLAMA 3.1), it is similarly documented by Meta in release reports that the same internal workflows were used, including for dataset curation (Grattafiori et al., 2024). This is generally considered definitive evidence that these models were also trained on `Books3` (as

well as additional books data from LibGen). (Our validity experiments enable us to independently conclude this. See Section 4 and Appendix F.)

2. **Category 2: DEEPSEEK V1 67B, QWEN 2.5 72B, and GEMMA 2 27B.** Although we do not know with certainty whether these models were trained on Books3, we aim to use our extraction methodology to assess whether they have memorized content from books. In particular, we want to design a measurement procedure that lets us conclude that reproducing such a sequence in an LLM's output is valid extraction (i.e., evidence of memorization). Based on our validity experiments (Section 4 and Appendix F), we can be very confident that when we successfully generate Books3 data, that those data were memorized from the training data. We discuss those experiments elsewhere, and note here that when we observe high probabilities p_z for these models, we can safely conclude that those sequences z have been memorized.
3. **Category 3: PHI 4.** For validity reasons, we also want to compare on a model that definitively should *not* have been trained on copyrighted books in Books3. (See also Appendix F.) We pick PHI 4 for these reasons, as it is a very high quality for its size class and Abdin et al. (2024) note in the PHI 4 report that it was trained predominantly on synthetic data. While we do not know which books are included in PHI 4, the report notes that it was not trained on whole copyrighted books. Public domain data may have been included. This makes PHI 4 and imperfect negative control, but we include it anyway as a basis for comparison. PHI 4 is similar in size to PYTHIA 12B, LLAMA 1 13B, and LLAMA 2 13B, which facilitates cross-model comparisons of extraction risk. (This is because the amount of memorization, reasonably, also varies according to model size; see also Carlini et al. (2023).)

Compute resources. We run all of our sliding-window experiments in a Slurm cluster environment, using the same node with 4 A100 GPUs. All told, this set of experiments took 394 GPU days.

Table 3: **Books drawn from Books3.** The 200 books from Books3 that we evaluate with our sliding-window probabilistic extraction procedure. See Sections 3 and 5, as well as Appendices E.1 & E.2.

	Author	Title	Year	Status	Random	Books3 path
1	Achebe, Chinua	*Things Fall Apart* (Achebe, 1958)	1958	©	False	https://the-eye.eu/public/Books/Bibliotik/C/Chinua%20Achebe%20-%20Things%20Fall%20Apart.epub
2	Adams, Douglas	*The Hitchhiker's Guide to the Galaxy - Omnibus* (Adams, 1979)	1979	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Hitchhiker%27s%20Guide%20to%20the%20Galaxy%20Omnibus%20A%20Trilogy%20in%20Five%20Parts%20-%20Douglas%20Adams.epub
3	Adichie, Chimamanda Ngozi	*Americanah* (Adichie, 2013)	2013	©	False	https://the-eye.eu/public/Books/Bibliotik/A/Americanah-Adichie.epub
4	Al Rawi, Shahad	*The Baghdad Clock* (Al Rawi, 2016)	2016	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Baghdad%20Clock%20-%20Shahad%20Al%20Rawi.epub
5	Armstrong, Kelley	*Industrial Magic* (Armstrong, 2004)	2004	©	True	https://the-eye.eu/public/Books/Bibliotik/I/Industrial%20Magic%20-%20Kelley%20Armstrong.epub
6	Asimov, Isaac	*Fantastic Voyage* (Asimov, 1966)	1966	©	False	https://the-eye.eu/public/Books/Bibliotik/I/IsaacAsimov-FantasticVoyage_9780307790965.epub
7	Asimov, Isaac	*The Complete Robot* (Asimov, 1982)	1982	©	False	https://the-eye.eu/public/Books/Bibliotik/I/Isaac%20Asimov%20-%20The%20Complete%20Robot.epub
8	Atwood, Margaret	*The Handmaid's Tale* (Atwood, 1985)	1985	©	False	https://the-eye.eu/public/Books/Bibliotik/M/Margaret%20Atwood%20%281985%29%20The%20Handmaid%27s%20Tale%20%5Bretail%5D.epub
9	Austen, Jane	*Pride and Prejudice* (Austen, 1813)	1813	PD	False	https://the-eye.eu/public/Books/Bibliotik/P/Pride%20and%20Prejudice%20-%20Jane%20Austen.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
10	Baldacci, David	*The Christmas Train* (Baldacci, 2001)	2001	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Christmas%20Train%20-%20David%20Baldacci.epub
11	Baldwin, James	*Notes of a Native Son* (Baldwin, 1955)	1955	©	False	https://the-eye.eu/public/Books/Bibliotik/J/James%20Baldwin%20-%20Notes%20of%20a%20Native%20Son.epub
12	Baldwin, James	*Another Country* (Baldwin, 1962)	1962	©	False	https://the-eye.eu/public/Books/Bibliotik/A/Another%20Country%20-%20James%20Baldwin.epub
13	Barnes, Julian	*The Lemon Table* (Barnes, 2004)	2004	©	True	https://the-eye.eu/public/Books/Bibliotik/J/Julian%20Barnes%20-%20The%20Lemon%20Table.epub
14	Barolini, Teodolinda	*Dante and the Origins of Italian Literary Culture* (Barolini, 2006)	2006	©	False	https://the-eye.eu/public/Books/Bibliotik/D/Dante%20and%20the%20Origins%20of%20Italia%20-%20Barolini%2C%20Teodolinda%3B.epub
15	Beard, Mary	*The Parthenon* (Beard, 2002)	2002	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Parthenon%20-%20Mary%20Beard.epub
16	Beardsley, Charles	*Guam: Past and Present* (Beardsley, 1964)	1964	©	True	https://the-eye.eu/public/Books/Bibliotik/G/Guam%20Past%20and%20Present%20-%20Charles%20Beardsley.epub
17	Beckett, Samuel	*Waiting for Godot* (Beckett, 1952)	1952	©	False	https://the-eye.eu/public/Books/Bibliotik/W/Waiting%20for%20Godot%20-%20Samuel%20Beckett.epub
18	Benedict, Helen	*The Lonely Soldier* (Benedict, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Lonely%20Soldier%202011%20-%20Helen%20Benedict.epub
19	Berry, Mary	*Simple Cakes* (Berry, 2006)	2006	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Simple%20Cakes%20-%20Mary%20Berry.epub
20	Box, C.J.	*Paradise Valley* (Box, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Paradise%20Valley%20-%20C.J.%20Box.epub
21	Bradbury, Ray	*The Cat's Pajamas* (Bradbury, 2004)	2004	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Cat%27s%20Pajamas%20-%20Ray%20Bradbury.epub
22	Bradshaw, Gillian	*London in Chains* (Bradshaw, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/L/London%20in%20Chains%20-%20Gillian%20Bradshaw.epub
23	Brockman, John	*My Einstein* (Brockman, 2006)	2006	©	True	https://the-eye.eu/public/Books/Bibliotik/M/My%20Einstein%20-%20John%20Brockman.epub
24	Brown, Dan	*The Da Vinci Code* (Brown, 2003)	2003	©	False	https://the-eye.eu/public/Books/Bibliotik/D/Dan%20Brown%20-%20The%20Da%20Vinci%20Code.epub
25	Bryant, Niobia	*Live and Learn* (Bryant, 2007)	2007	©	True	https://the-eye.eu/public/Books/Bibliotik/L/Live%20and%20Learn%20-%20Niobia%20Bryant.epub
26	Brzezinski, Mika	*Knowing Your Value* (Brzezinski, 2011)	2011	©	True	https://the-eye.eu/public/Books/Bibliotik/K/Knowing%20Your%20Value%20-%20Mika%20Brzezinski.epub
27	Camus, Albert	*The Myth of Sisyphus* (Camus, 1942)	1942	©	False	https://the-eye.eu/public/Books/Bibliotik/2/2013%28orig1942%29%20Albert%20Camus%20-%20The%20Myth%20of%20Sisyphus%5BTransl%20Justin%20O%27Brien%5D_Ral.epub
28	Carroll, Lewis	*Alice's Adventures in Wonderland and Through the Looking-Glass and What Alice Found There* (Carroll, 1871)	1871	PD	False	https://the-eye.eu/public/Books/Bibliotik/A/Alice%27s%20Adventures%20in%20Wonderlan%20-%20Lewis%20Carroll.epub
29	Carver, Jeffrey A.	*The Infinity Link* (Carver, 1984)	1984	©	True	https://the-eye.eu/public/Books/Bibliotik/I/Infinity%20Link%20-%20Jeffrey%20A.%20Carver.epub
30	Christie, Agatha	*Murder on the Orient Express* (Christie, 1933)	1933	©	False	https://the-eye.eu/public/Books/Bibliotik/M/Murder%20on%20the%20Orient%20Express-Christie.epub
31	Christie, Agatha	*And Then There Were None* (Christie, 1939)	1939	©	True	https://the-eye.eu/public/Books/Bibliotik/A/And%20Then%20There%20Were%20None%20-%20Agatha%20Christie.epub
32	Coates, Ta-Nehisi	*The Beautiful Struggle* (Coates, 2008)	2008	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Beautiful%20Struggle%20-%20Ta-Nehisi%20Coates.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
33	Coates, Ta-Nehisi	*We Were Eight Years in Power* (Coates, 2017)	2017	©	False	https://the-eye.eu/public/Books/Bibliotik/W/We%20Were%20Eight%20Years%20in%20Power%20-%20Ta-Nehisi%20Coates.epub
34	Coates, Ta-Nehisi	*The Water Dancer* (Coates, 2019)	2019	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Water%20Dancer%20-%20Ta-Nehisi%20Coates%3B.epub
35	Cocteau, Jean	*The Infernal Machine and Other Plays* (Cocteau, 1963)	1963	©	True	https://the-eye.eu/public/Books/Bibliotik/I/Infernal%20Machine%2C%20The%20-%20Jean%20Cocteau.epub
36	Coelho, Paulo	*The Alchemist* (Coelho, 1988)	1988	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Alchemist%20%2825th%20Anniversary%20Edition%29%20-%20Paulo%20Coelho.epub
37	Cogburn, Jon and Silcox, Mark	*Dungeons and Dragons and Philosophy* (Cogburn & Silcox, 2012)	2012	©	True	https://the-eye.eu/public/Books/Bibliotik/D/Dungeons%20and%20Dragons%20and%20Philosophy%20-%20Christopher%20Robichaud.epub
38	Cohen-Solal, Annie	*Mark Rothko* (Cohen-Solal, 2013)	2013	©	True	https://the-eye.eu/public/Books/Bibliotik/M/Mark%20Rothko%20-%20Annie%20Cohen-Solal.epub
39	Collins, Suzanne	*The Hunger Games* (Collins, 2008)	2008	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Hunger%20Games%20-%20Suzanne%20Collins.epub
40	Cook, Glen	*The Dragon Never Sleeps* (Cook, 1988)	1988	©	True	https://the-eye.eu/public/Books/Bibliotik/G/Glen%20Cook%20-%20The%20Dragon%20Never%20Sleeps%20%5Bretail%5D.epub
41	Covey, Stephen R.	*The 7 Habits of Highly Effective People* (Covey, 1989)	1989	©	False	https://the-eye.eu/public/Books/Bibliotik/S/Stephen%20Covey%20-%20The%207%20Habits%20of%20Highly%20Effective%20People.epub
42	Crabb, David	*Bad Kid* (Crabb, 2015)	2015	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Bad%20Kid%20-%20David%20Crabb.epub
43	Crais, Robert	*Lullaby Town* (Crais, 1992)	1992	©	True	https://the-eye.eu/public/Books/Bibliotik/L/Lullaby%20Town%20-%20Robert%20Crais.epub
44	Crichton, Michael	*Jurassic Park* (Crichton, 1990)	1990	©	False	https://the-eye.eu/public/Books/Bibliotik/J/Jurassic_Park-Crichton.epub
45	Cunningham, Michael	*The Hours* (Cunningham, 1998)	1998	©	True	https://the-eye.eu/public/Books/Bibliotik/M/Michael%20Cunningham%20-%20The%20Hours.epub
46	Czapski, Józef	*Inhuman Land* (Czapski, 1949)	1949	©	True	https://the-eye.eu/public/Books/Bibliotik/I/Inhuman%20Land%20-%20Jozef%20Czapski.epub
47	Dahl, Roald	*Charlie and the Chocolate Factory* (Dahl, 1964)	1964	©	False	https://the-eye.eu/public/Books/Bibliotik/C/Charlie%20and%20the%20Chocolate%20Factory%20-%20Roald%20Dahl.epub
48	Dahl, Roald	*James and the Giant Peach* (Dahl, 1961)	1961	©	True	https://the-eye.eu/public/Books/Bibliotik/R/Roald%20Dahl%20-%20James%20and%20the%20Giant%20Peach-%20A%20Play%20%28retail%29.epub
49	Diakopoulos, Nicholas	*Automating the News* (Diakopoulos, 2019)	2019	©	True	https://the-eye.eu/public/Books/Bibliotik/N/Nicholas%20Diakopoulos%20-%20Automating%20the%20News.epub
50	Díaz, Junot	*Drown* (Díaz, 1995)	1995	©	False	https://the-eye.eu/public/Books/Bibliotik/J/Junot%20Diaz%20-%20Drown.epub
51	Díaz, Junot	*The Brief Wondrous Life of Oscar Wao* (Díaz, 2007)	2007	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Brief%20Wondrous%20Life%20of%20Oscar%20Wao.epub
52	Díaz, Junot	*This Is How You Lose Her* (Díaz, 2012)	2012	©	False	https://the-eye.eu/public/Books/Bibliotik/T/This%20Is%20How%20You%20Lose%20Her-Diaz.epub
53	Didion, Joan	*The White Album* (Didion, 1979)	1979	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20White%20Album%20%28Zola%20Books%29%20-%20Joan%20Didion%20%28retail%29.epub
54	Doctorow, Cory	*Down and Out in the Magic Kingdom* (Doctorow, 2003)	2003	CC-BY-SA	False	https://the-eye.eu/public/Books/Bibliotik/D/Down%20and%20Out%20in%20the%20Magic%20Kingd%20-%20Cory%20Doctorow.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
55	Duffy, Carol Ann	*The World's Wife* (Duffy, 1999)	1999	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20World%27s%20Wife%20-%20Carol%20Ann%20Duffy.epub
56	Egan, Jennifer	*A Visit from the Goon Squad* (Egan, 2010)	2010	©	False	https://the-eye.eu/public/Books/Bibliotik/A/A%20Visit%20from%20the%20Goon%20Squad%20-%20Jennifer%20Egan.epub
57	Ellison, Ralph	*Invisible Man* (Ellison, 1952)	1952	©	False	https://the-eye.eu/public/Books/Bibliotik/I/Invisible%20Man%20-%20Ralph%20Ellison.epub
58	Fairless, Chelsea and Garroni, Lauren	*We Should All Be Mirandas* (Fairless & Garroni, 2019)	2019	©	True	https://the-eye.eu/public/Books/Bibliotik/W/We%20Should%20All%20Be%20Mirandas%20-%20Chelsea%20Fairless.epub
59	Falsani, Cathleen	*The Dude Abides* (Falsani, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Dude%20Abides%20-%20Cathleen%20Falsani.epub
60	Farnsworth, Christopher	*The President's Vampire* (Farnsworth, 2011)	2011	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20President%27s%20Vampire%20-%20Christopher%20Farnsworth.epub
61	Fitzgerald, F. Scott	*The Great Gatsby* (Fitzgerald, 1925)	1925	PD	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Great%20Gatsby%20-%20F.%20Scott%20Fitzgerald.epub
62	Flynn, Gillian	*Gone Girl* (Flynn, 2012)	2012	©	False	https://the-eye.eu/public/Books/Bibliotik/G/Gone%20Girl%20-%20Gillian%20Flynn.epub
63	Friedman, Norman	*British Destroyers* (Friedman, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/B/British%20Destroyers%20-%20Norman%20Friedman.epub
64	Fudurich, Barbara and Hansen, Joan and Caroline Linscott and Medway, Geri	*Florals & Botanicals* (Fudurich et al., 2008)	2008	©	True	https://the-eye.eu/public/Books/Bibliotik/F/Florals%20%26%20Botanicals%20-%20Barbara%20Fudurich.epub
65	Gaiman, Neil and Pratchett, Terry	*Good Omens* (Gaiman & Pratchett, 1990)	1990	©	False	https://the-eye.eu/public/Books/Bibliotik/T/Terry%20Pratchett%2C%20Neil%20Gaiman%20-%20Good%20Omens%20%5BUK%5D%20%5Bretail%5D.epub
66	Gideon, Melanie	*The Slippery Year* (Gideon, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Slippery%20Year%20-%20Melanie%20Gideon.epub
67	Girard, Stefanie	*Sweater Surgery* (Girard, 2010)	2010	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Sweater%20Surgery%20-%20Stefanie%20Girard.epub
68	Gladwell, Malcolm	*Blink* (Gladwell, 2005)	2005	©	False	https://the-eye.eu/public/Books/Bibliotik/B/Blink%20-%20Malcolm%20Gladwell.epub
69	Glover, Richard	*The Land Before Avocado* (Glover, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/R/Richard%20Glover%20-%20The%20Land%20Before%20Avocado.epub
70	Golden, Christopher	*Dead Ringers* (Golden, 2015)	2015	©	False	https://the-eye.eu/public/Books/Bibliotik/C/Christopher%20Golden%20-%20Dead%20Ringers%20%5Bretail%5D.epub
71	Golden, Christopher	*Ararat* (Golden, 2017)	2017	©	False	https://the-eye.eu/public/Books/Bibliotik/A/Ararat%20-%20Christopher%20Golden.epub
72	Golding, William	*Lord of the Flies* (Golding, 1954)	1954	©	False	https://the-eye.eu/public/Books/Bibliotik/L/Lord%20of%20the%20Flies%20-%20William%20Golding.epub
73	Goodkind, Terry	*Wizard's First Rule* (Goodkind, 1994)	1994	©	True	https://the-eye.eu/public/Books/Bibliotik/W/Wizard%27s%20First%20Rule%200101%20-%20Terry%20Goodkind.epub
74	Goodman, Martin	*Rome and Jerusalem* (Goodman, 2007)	2007	©	True	https://the-eye.eu/public/Books/Bibliotik/R/Rome%20and%20Jerusalem%20-%20Martin%20Goodman.epub
75	Green, John	*The Fault in Our Stars* (Green, 2012)	2012	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Fault%20in%20Our%20Stars%20-%20John%20Green.epub
76	Greene, Graham	*The Third Man* (Greene, 1950)	1950	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Third%20Man%20-%20Graham%20Greene.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
77	Greer, Andrew Sean	*The Confessions of Max Tivoli* (Greer, 2004)	2004	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Confessions%20of%20Max%20Tivoli%20-%20Andrew%20Sean%20Greer.epub
78	Grisham, John	*The Fugitive* (Grisham, 2015)	2015	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Fugitive%20-%20John%20Grisham.epub
79	Haddon, Mark	*The Curious Incident of the Dog in the Night-Time* (Haddon, 2003)	2003	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Curious%20Incident%20of%20the%20Dog%20in%20the%20Night-Time%20-%20Mark%20Haddon.epub
80	Halpern, Sue	*Migrations to Solitude* (Halpern, 1992)	1992	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Sue%20Halpern%20-%20Migrations%20to%20Solitude.epub
81	Hanks, Tom	*Uncommon Type* (Hanks, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/U/Uncommon%20Type%20-%20Tom%20Hanks.epub
82	Hanson, Thor	*Buzz* (Hanson, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Buzz%20-%20Thor%20Hanson.epub
83	Haydon, Elizabeth	*Requiem for the Sun* (Haydon, 2002)	2002	©	True	https://the-eye.eu/public/Books/Bibliotik/R/Requiem%20for%20the%20Sun%20-%20Elizabeth%20Haydon.epub
84	Heller, Joseph	*Catch-22* (Heller, 1961)	1961	©	False	https://the-eye.eu/public/Books/Bibliotik/C/Catch-22%20-%20Joseph%20Heller.epub
85	Hemingway, Ernest	*The Old Man and the Sea* (Hemingway, 1952)	1952	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Old%20Man%20and%20the%20Sea%20-%20Ernest%20Hemingway.epub
86	Hendy, David	*Life on Air* (Hendy, 2007)	2007	©	True	https://the-eye.eu/public/Books/Bibliotik/L/Life%20on%20Air%20-%20Hendy%2C%20David%3B.epub
87	Hersheson, Luke	*Great Hair Days* (Hersheson, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/G/Great%20Hair%20Days%20-%20Luke%20Hersheson.epub
88	Hill, Matt	*Graft* (Hill, 2016)	2016	©	True	https://the-eye.eu/public/Books/Bibliotik/G/Graft%20-%20Matt%20Hill.epub
89	Hinton, S. E.	*The Outsiders* (Hinton, 1967)	1967	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Outsiders%20-%20SE%20Hinton.epub
90	Huff, Tanya	*The Second Summoning* (Huff, 2001)	2001	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Second%20Summoning%20-%20Tanya%20Huff.epub
91	Hughes, Langston	*Selected Poems of Langston Hughes* (Hughes, 1959)	1959	©	False	https://the-eye.eu/public/Books/Bibliotik/S/Selected%20Poems%20of%20Langston%20Hugh%20-%20Langston%20Hughes.epub
92	Hwang, David Henry	*M. Butterfly* (Hwang, 1988)	1988	©	False	https://the-eye.eu/public/Books/Bibliotik/M/M.%20Butterfly%20-%20David%20Henry%20Hwang.epub
93	Jackson, Allen	*Building and Operating a Realistic Model Railway* (Jackson, 2016)	2016	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Building%20and%20Operating%20a%20Realistic%20Model%20Railway%20-%20Allen%20Jackson.epub
94	Jacobs, Betty	*All the Onions* (Jacobs, 1993)	1993	©	True	https://the-eye.eu/public/Books/Bibliotik/A/All%20the%20Onions%20-%20Betty%20E.%20M.%20Jacobs.epub
95	James, E. L.	*Fifty Shades of Grey* (James, 2011)	2011	©	False	https://the-eye.eu/public/Books/Bibliotik/F/Fifty%20Shades%20Darker%20-%20E%20L%20James.epub
96	Jemisin, N. K.	*The Stone Sky* (Jemisin, 2017)	2017	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Stone%20Sky%20-%20N.%20K.%20Jemisin.epub
97	Joyce, James	*Ulysses* (Joyce, 1922)	1922	PD	False	https://the-eye.eu/public/Books/Bibliotik/U/Ulysses%20-%20James%20Joyce%20-%20Penguin%20Group%20-%202000.epub
98	Kadrey, Richard	*Sandman Slim* (Kadrey, 2009)	2009	©	False	https://the-eye.eu/public/Books/Bibliotik/S/Sandman%20Slim%20-%20Richard%20Kadrey.epub
99	Kaldellis, Anthony	*Ethnography after Antiquity* (Kaldellis, 2013)	2013	©	True	https://the-eye.eu/public/Books/Bibliotik/E/Ethnography%20After%20Antiquity%20-%20Anthony%20Kaldellis.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
100	Klam, Matthew	*Who Is Rich?* (Klam, 2017)	2017	©	False	https://the-eye.eu/public/Books/Bibliotik/W/Who%20Is%20Rich_%20-%20Matthew%20Klam.epub
101	Koontz, Dean	*The Servants of Twilight* (Koontz, 1984)	1984	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Servants%20of%20Twilight%20-%20Dean%20Koontz.epub
102	Kuhn, Aihan	*Tai Chi for Depression* (Kuhn, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/T/Tai%20Chi%20for%20Depression%20-%20Aihan%20Kuhn.epub
103	Kun, Josh	*The Tide Was Always High* (Kun, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Tide%20Was%20Always%20High%20-%20Josh%20Kun.epub
104	Kwok, Jean	*Girl in Translation* (Kwok, 2010)	2010	©	True	https://the-eye.eu/public/Books/Bibliotik/J/JeanKwok-GirlinTranslation_9781101182673.epub
105	L'Engle, Madeleine	*A Wrinkle in Time* (L'Engle, 1962)	1962	©	False	https://the-eye.eu/public/Books/Bibliotik/A/A%20Wrinkle%20in%20Time%20-%20Madeleine%20L%27engle.epub
106	Lago, Eduardo	*Call Me Brooklyn* (Lago, 2006)	2006	©	True	https://the-eye.eu/public/Books/Bibliotik/C/Call%20Me%20Brooklyn%20-%20Eduardo%20Lago.epub
107	Larson, Erik	*Dead Wake* (Larson, 2015)	2015	©	True	https://the-eye.eu/public/Books/Bibliotik/D/Dead%20Wake%20-%20Erik%20Larson.epub
108	Larsson, Stieg	*The Girl with the Dragon Tattoo* (Larsson, 2005)	2005	©	False	https://the-eye.eu/public/Books/Bibliotik/G/Girl%20with%20the%20Dragon%20Tattoo%20-%20Stieg%20Larsson.epub
109	Le Guin, Ursula K.	*The Daughter of Odren* (Le Guin, 2014)	2014	©	False	https://the-eye.eu/public/Books/Bibliotik/U/Ursula%20K.%20Le%20Guin%20-%20Earthsea%2006.5%20-%20The%20Daughter%20of%20Odren%20%5Bretail%5D.epub
110	Lewis, C. S.	*The Chronicles of Narnia* (Lewis, 1950)	1950	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Chronicles%20of%20Narnia%20-%20C.S.%20Lewis.epub
111	Lippman, Laura	*After I'm Gone* (Lippman, 2014)	2014	©	False	https://the-eye.eu/public/Books/Bibliotik/A/After%20I%27m%20Gone%20-%20Laura%20Lippman.epub
112	Lippman, Laura	*Sunburn* (Lippman, 2018)	2018	©	False	https://the-eye.eu/public/Books/Bibliotik/S/Sunburn%20-%20Laura%20Lippman.epub
113	Liss, David	*Marvel's Spider-Man: Hostile Takeover* (Liss, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/M/Marvel%27s%20Spider-Man_%20Hostile%20Ta%20-%20David%20Liss.epub
114	Lowry, Lois	*Anastasia on Her Own* (Lowry, 1985)	1985	©	False	https://the-eye.eu/public/Books/Bibliotik/A/Anastasia%20on%20Her%20Own%20-%20Lois%20Lowry.epub
115	MacMillan, Margaret	*Nixon in China* (MacMillan, 2006)	2006	©	True	https://the-eye.eu/public/Books/Bibliotik/N/Nixon%20in%20China%20-%20Margaret%20Macmillan.epub
116	Mariani, Scott	*The Doomsday Prophecy* (Mariani, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Doomsday%20Prophecy%20-%20Scott%20Mariani.epub
117	Martin, George R. R.	*A Game of Thrones* (Martin, 1996)	1996	©	False	https://the-eye.eu/public/Books/Bibliotik/A/A%20Game%20of%20Thrones%20-%20George%20R.%20R.%20Martin.epub
118	Meldahl, Keith Heyer	*Rough-Hewn Land* (Meldahl, 2011)	2011	©	True	https://the-eye.eu/public/Books/Bibliotik/R/Rough-Hewn%20Land.epub
119	Meyer, Stephenie	*Twilight* (Meyer, 2005)	2005	©	False	https://the-eye.eu/public/Books/Bibliotik/T/Twilight%20-%20Stephenie%20Meyer.epub
120	Milan, Courtney	*The Duchess War* (Milan, 2012)	2012	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Duchess%20War%20-%20Courtney%20Milan.epub
121	Mill, John Stuart	*On Liberty* (Mill, 1859)	1859	PD	False	https://the-eye.eu/public/Books/Bibliotik/O/On%20Liberty%20-%20John%20Stuart%20Mill.epub
122	Miller, Alex	*Coal Creek* (Miller, 2013)	2013	©	True	https://the-eye.eu/public/Books/Bibliotik/C/Coal%20Creek%20-%20Alex%20Miller.epub
123	Milne, A. A.	*Winnie the Pooh* (Milne, 1926)	1926	PD	False	https://the-eye.eu/public/Books/Bibliotik/W/Winnie_the_Pooh-Milne.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
124	Moore, Colten and O'Brien, Keith	*Catching the Sky* (Moore & O'Brien, 2016)	2016	©	True	https://the-eye.eu/public/Books/Bibliotik/C/Catching%20the%20Sky%20-%20Colten%20Moore.epub
125	Moran, Michelle	*The Heretic Queen* (Moran, 2008)	2008	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Heretic%20Queen%20-%20Michelle%20Moran.epub
126	Morelli, Laura	*The Gondola Maker* (Morelli, 2014)	2014	©	True	https://the-eye.eu/public/Books/Bibliotik/G/Gondola%20Maker%2C%20The%20-%20Laura%20Morelli.epub
127	Morris, Mary McGarry	*Songs in Ordinary Time* (Morris, 1995)	1995	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Songs%20in%20Ordinary%20Time%20-%20Mary%20McGarry%20Morris.epub
128	Morrison, Toni	*Beloved* (Morrison, 1987a)	1987	©	False	https://the-eye.eu/public/Books/Bibliotik/B/Beloved%20-%20Toni%20Morrison.epub
129	Murakami, Haruki	*Norwegian Wood* (Murakami, 1987)	1987	©	False	https://the-eye.eu/public/Books/Bibliotik/N/Norwegian%20Wood%20%28Vintage%20Books%29%20%5BEpub%5D%20-%20Haruki%20Murakami.epub
130	Nielsen, Desiree	*Eat More Plants* (Nielsen, 2019)	2019	©	True	https://the-eye.eu/public/Books/Bibliotik/E/Eat%20More%20Plants%20-%20Desiree%20Nielsen%3B.epub
131	Northrop, Michael	*Polaris* (Northrop, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Polaris%20-%20Michael%20Northrop.epub
132	O'Donnell, James J.	*Pagans* (O'Donnell, 2015)	2015	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Pagans%20-%20James%20J.%20O%27Donnell.epub
133	O'Sullivan, Meghan L.	*Windfall* (O'Sullivan, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/W/Windfall%20-%20Meghan%20L.%20O%27Sullivan.epub
134	Ogawa, Yoko	*The Memory Police* (Ogawa, 1994)	1994	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Memory%20Police%20-%20Yoko%20Ogawa.epub
135	Oliveira, Robin	*Winter Sisters* (Oliveira, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/W/Winter%20Sisters%20-%20Robin%20Oliveira.epub
136	Orwell, George	*Nineteen Eighty-Four* (Orwell, 1949)	1949	©	False	https://the-eye.eu/public/Books/Bibliotik/N/Nineteen%20Eighty-Four%20%28The%20Annotated%20Edition%29%20-%20George%20Orwell.epub
137	Palahniuk, Chuck	*Fight Club* (Palahniuk, 1996)	1996	©	False	https://the-eye.eu/public/Books/Bibliotik/F/Fight%20Club%20-%20Chuck%20Palahniuk.epub
138	Papazian, Charlie	*The Complete Joy of Homebrewing* (Papazian, 1984)	1984	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Complete%20Joy%20of%20Homebrewing%20-%20Charlie%20Papazian.epub
139	Park, Paul	*The Cult of Loving Kindness* (Park, 1991)	1991	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Cult%20of%20Loving%20Kindness%20-%20Paul%20Park.epub
140	Patterson, James	*Along Came a Spider* (Patterson, 1993)	1993	©	False	https://the-eye.eu/public/Books/Bibliotik/A/Along%20Came%20a%20Spider%20-%20James%20Patterson.epub
141	Payard, François and McBride, Anne E.	*Payard Cookies* (Payard & McBride, 2015)	2015	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Payard%20Cookies.epub
142	Pépin, Jacques	*Essential Pepin Desserts* (Pépin, 2011)	2011	©	True	https://the-eye.eu/public/Books/Bibliotik/E/Essential%20Pepin%20Desserts%20-%20Jacques%20Pepin.epub
143	Piazza, Tom	*Why New Orleans Matters* (Piazza, 2005)	2005	©	True	https://the-eye.eu/public/Books/Bibliotik/W/Why%20New%20Orleans%20Matters%20-%20Tom%20Piazza.epub
144	Pinker, Steven	*Enlightenment Now* (Pinker, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/E/Enlightenment%20Now%20-%20Steven%20Pinker.epub
145	Porter, Michael E.	*Competitive Strategy* (Porter, 1980)	1980	©	True	https://the-eye.eu/public/Books/Bibliotik/C/Competitive%20Strategy%200101%20-%20Michael%20E.%20Porter.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
146	Pratchett, Terry	*Night Watch* (Pratchett, 2002)	2002	©	False	https://the-eye.eu/public/Books/Bibliotik/T/Terry%20Pratchett%20-%20Discworld%2029%20-%20Night%20Watch%20%5Bretail%5D.epub
147	Pullman, Philip	*The Subtle Knife* (Pullman, 1997)	1997	©	False	https://the-eye.eu/public/Books/Bibliotik/P/Pullman%20-%20The%20Subtle%20Knife.epub
148	Rampley, Matthew	*The Seductions of Darwin* (Rampley, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Seductions%20of%20Darwin%20-%20Matthew%20Rampley.epub
149	Remen, Rachel Naomi	*Kitchen Table Wisdom* (Remen, 1996)	1996	©	True	https://the-eye.eu/public/Books/Bibliotik/K/Kitchen%20Table%20Wisdom%2010th%20Anniversary%20-%20Rachel%20Naomi%20Remen.epub
150	Roberts, Jessi and Witter, Bret	*Backroads Boss Lady* (Roberts & Witter, 2019)	2019	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Backroads%20Boss%20Lady%20-%20Jessi%20Roberts.epub
151	Roger, Marie-Sabine	*Soft in the Head* (Roger, 2008)	2008	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Soft%20in%20the%20Head%20-%20Marie-Sabine%20Roger.epub
152	Rouighi, Ramzi	*The Making of a Mediterranean Emirate* (Rouighi, 2011)	2011	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Making%20of%20a%20Mediterranean%20E%20-%20Rouighi%2C%20Ramzi%3B.epub
153	Rowling, J. K.	*Harry Potter and the Sorcerer's Stone* (Rowling, 1997)	1997	©	False	https://the-eye.eu/public/Books/Bibliotik/H/Harry_Potter_and_the_Sorcerers_Stone-Rowling.epub
154	Rowling, J. K.	*Harry Potter and the Chamber of Secrets* (Rowling, 1998)	1998	©	False	https://the-eye.eu/public/Books/Bibliotik/H/Harry_Potter_and_the_Chamber_of_Secrets-Rowling.epub
155	Rowling, J. K.	*Harry Potter and the Goblet of Fire* (Rowling, 2000)	2000	©	False	https://the-eye.eu/public/Books/Bibliotik/H/Harry_Potter_and_the_Goblet_of_Fire-Rowling.epub
156	Rowling, J. K.	*Harry Potter and the Deathly Hallows* (Rowling, 2007)	2007	©	False	https://the-eye.eu/public/Books/Bibliotik/H/Harry_Potter_and_the_Deathly_Hallows-Rowling.epub
157	Rubinstein, Dan	*Born to Walk* (Rubinstein, 2015)	2015	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Born%20to%20Walk_%20The%20Transformative%20Power%20of%20a%20Pedestrian%20Act%20-%20Dan%20Rubinstein%20%28retail%29.epub
158	Ruskin, Marc	*The Pretender* (Ruskin, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Pretender%20-%20Marc%20Ruskin.epub
159	Sachs, Harvey	*Toscanini* (Sachs, 1978)	1978	©	True	https://the-eye.eu/public/Books/Bibliotik/T/Toscanini%20-%20Harvey%20Sachs.epub
160	Sagan, Carl	*Cosmos* (Sagan, 1980)	1980	©	False	https://the-eye.eu/public/Books/Bibliotik/C/Cosmos%20-%20Carl%20Sagan.epub
161	Sahni, Bhisham	*Middle India* (Sahni, 2001)	2001	©	True	https://the-eye.eu/public/Books/Bibliotik/M/Middle%20India%20%28Penguin%29%20-%20Bhisam%20Sahni.epub
162	Salinger, J. D.	*The Catcher in the Rye* (Salinger, 1951)	1951	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Catcher%20in%20the%20Rye%20-%20J.%20D.%20Salinger.epub
163	Sandberg, Sheryl	*Lean in* (Sandberg, 2013)	2013	©	False	https://the-eye.eu/public/Books/Bibliotik/L/Lean%20In%20-%20Sheryl%20Sandberg.epub
164	Sharma, Sanjeev	*The DevOps Adoption Playbook* (Sharma, 2017)	2017	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20DevOps%20Adoption%20Playbook%202017%20-%20Sharma%2C%20Sanjeev%3B.epub
165	Shelley, Mary	*Frankenstein* (Shelley, 1818)	1818	PD	False	https://the-eye.eu/public/Books/Bibliotik/F/Frankenstein%20-%20Mary%20Shelley.epub
166	Sherr, Lynn	*Sally Ride: America's First Woman in Space* (Sherr, 2014)	2014	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Sally%20Ride_%20America%27s%20First%20Woman%20in%20Space%20-%20Lynn%20Sherr.epub
167	Shriver, Lionel	*A Perfectly Good Family* (Shriver, 1996)	1996	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Perfectly%20Good%20Family%2C%20A%20-%20Lionel%20Shriver.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
168	Silverman, Sarah	*The Bedwetter* (Silverman, 2010)	2010	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Bedwetter%20-%20Sarah%20Silverman.epub
169	Sloman, Larry	*On the Road with Bob Dylan* (Sloman, 1978)	1978	©	True	https://the-eye.eu/public/Books/Bibliotik/O/On%20the%20Road%20with%20Bob%20Dylan%20-%20Larry%20Sloman.epub
170	Smith, Alexander Gordon	*The Night Children* (Smith, 2008)	2008	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Night%20Children%20-%20Alexander%20Gordon%20Smith.epub
171	Smith, Zadie	*White Teeth* (Smith, 2000)	2000	©	False	https://the-eye.eu/public/Books/Bibliotik/S/Smith%2C%20Zadie%20-%20White%20Teeth%20%28ISBN%209781400075508%29.epub
172	Snyder, Rachel Louise	*No Visible Bruises* (Snyder, 2019)	2019	©	False	https://the-eye.eu/public/Books/Bibliotik/N/No%20Visible%20Bruises%20-%20Rachel%20Louise%20Snyder.epub
173	Steinbeck, John	*The Grapes of Wrath* (Steinbeck, 1939)	1939	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Grapes%20of%20Wrath%20-%20John%20Steinbeck.epub
174	Stiles, T. J.	*Jesse James* (Stiles, 2002)	2002	©	True	https://the-eye.eu/public/Books/Bibliotik/J/Jesse%20James%20-%20T.J.%20Stiles.epub
175	Stine, R. L.	*Zombie Halloween* (Stine, 2014)	2014	©	False	https://the-eye.eu/public/Books/Bibliotik/Z/Zombie%20Halloween%20-%20R.L.%20Stine.epub
176	Stubbs, David	*Fear of Music* (Stubbs, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/D/David%20Stubbs%20-%20Fear%20of%20Music.epub
177	Swanson, Jennifer	*Pearl Harbor* (Swanson, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Pearl%20Harbor%20-%20Jennifer%20Swanson.epub
178	Swift, Earl	*Chesapeake Requiem* (Swift, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/C/Chesapeake%20Requiem%20-%20Earl%20Swift.epub
179	Tartt, Donna	*The Goldfinch* (Tartt, 2013)	2013	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Goldfinch-Tartt.epub
180	TerKeurst, Lysa	*Unglued* (TerKeurst, 2012)	2012	©	False	https://the-eye.eu/public/Books/Bibliotik/L/Lysa%20TerKeurst%2C%20Unglued.epub
181	TerKeurst, Lysa	*Embraced* (TerKeurst, 2018)	2018	©	False	https://the-eye.eu/public/Books/Bibliotik/E/Embraced%20-%20Lysa%20Terkeurst.epub
182	Thomson, Lynn	*Birding with Yeats* (Thomson, 2014)	2014	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Birding%20with%20Yeats%20-%20Lynn%20Thomson.epub
183	Tolkien, J. R. R.	*The Hobbit* (Tolkien, 1937)	1937	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Hobbit%20%28Houghton%20Mifflin%20Harcourt%29%20%2875th%20Anniversary%20Edition%29%20%5BEpub%5D%20-%20J.R.R.%20Tolkien.epub
184	Tolkien, J. R. R.	*The Fellowship of the Ring* (Tolkien, 1954)	1954	©	False	https://the-eye.eu/public/Books/Bibliotik/J/J._R._R._Tolkien_-_The_Fellowship_of_the_Ring_Being_the_First_Part_of_The_Lord_of_the_Rings_-_Houghton_Mifflin_Harcourt_-_v4.0615_-_2012_-_RETAIL.epub
185	Tolkien, J. R. R.	*Tree and Leaf* (Tolkien, 1964)	1964	©	False	https://the-eye.eu/public/Books/Bibliotik/T/Tree%20and%20Leaf%20-%20J.%20R.%20R.%20Tolkien.epub
186	Trang, Corinne	*Noodles Every Day* (Trang, 2009)	2009	©	True	https://the-eye.eu/public/Books/Bibliotik/N/Noodles%20Every%20Day%20-%20Maura%20McEvoy.epub
187	Tyler, George R.	*Billionaire Democracy* (Tyler, 2018)	2018	©	True	https://the-eye.eu/public/Books/Bibliotik/B/Billionaire%20Democracy%20-%20George%20R.%20Tyler.epub
188	Venter, Al J.	*Portugal's Guerrilla Wars in Africa* (Venter, 2013)	2013	©	True	https://the-eye.eu/public/Books/Bibliotik/P/Portugal%27s%20Guerrilla%20Wars%20in%20Africa%20-%20Al%20Venter.epub

Continued on next page

Table 3 (continued). The 200 books drawn from Books3.

	Author	Title	Year	Status	Random	Books3 path
189	Vonnegut, Kurt	*Slaughterhouse-Five* (Vonnegut, 1969)	1969	©	True	https://the-eye.eu/public/Books/Bibliotik/S/Slaughterhouse-Five%20-%20Kurt%20Vonnegut%20%28retail%29.epub
190	Waldau, Paul	*Animal Rights* (Waldau, 2011)	2011	©	True	https://the-eye.eu/public/Books/Bibliotik/A/Animal%20Rights%20-%20Waldau%2C%20Paul.epub
191	Ward, Jesmyn	*Men We Reaped* (Ward, 2013)	2013	©	False	https://the-eye.eu/public/Books/Bibliotik/M/Men%20We%20Reaped%20-%20Jesmyn%20Ward.epub
192	White, E. B.	*Charlotte's Web* (White, 1952)	1952	©	False	https://the-eye.eu/public/Books/Bibliotik/C/Charlotte%27s%20Web%20-%20E.%20B.%20White.epub
193	Williamson, Marianne	*A Return to Love* (Williamson, 1992)	1992	©	False	https://the-eye.eu/public/Books/Bibliotik/M/Marianne%20Williamson%20-%20A%20Return%20to%20Love.epub
194	Woodson, Jacqueline	*Another Brooklyn* (Woodson, 2016)	2016	©	False	https://the-eye.eu/public/Books/Bibliotik/A/Another%20Brooklyn%20-%20Jacqueline%20Woodson.epub
195	Woodson, Jacqueline	*Brown Girl Dreaming* (Woodson, 2014)	2014	©	False	https://the-eye.eu/public/Books/Bibliotik/B/Brown%20Girl%20Dreaming%20-%20Jacqueline%20Woodson.epub
196	Yanagihara, Hanya	*A Little Life* (Yanagihara, 2015)	2015	©	False	https://the-eye.eu/public/Books/Bibliotik/A/A%20Little%20Life%20-%20Hanya%20Yanagihara.epub
197	Yoshimura, Yuji and Halford, Giovanna M.	*The Art of Bonsai* (Yoshimura & Halford, 1996)	1996	©	True	https://the-eye.eu/public/Books/Bibliotik/T/The%20Art%20of%20Bonsai%20-%20Yuji%20Yoshimura.epub
198	Zinn, Howard	*A People's History of the United States* (Zinn, 1980)	1980	©	False	https://the-eye.eu/public/Books/Bibliotik/A/A_Peoples_History_of_the_United_States-Zinn.epub
199	Zittrain, Jonathan	*The Future of the Internet and How to Stop It* (Zittrain, 2008)	2008	©	False	https://the-eye.eu/public/Books/Bibliotik/J/Jonathan%20Zittrain%20-%20The%20Future%20of%20the%20Internet.epub
200	Zusak, Markus	*The Book Thief* (Zusak, 2005)	2005	©	False	https://the-eye.eu/public/Books/Bibliotik/T/The%20Book%20Thief%20-%20Markus%20Zusak.epub

E.2 Sliding-window probabilistic extraction procedure

In this appendix, we provide more details on our main extraction experiments (Section 3). These experiments take the approach of constructing 100-token sequences z out of Books3 books, which we attempt to extract. These sequences are constructed using a sliding window across the length of an entire book, and we run these experiments for each of the 200 books in Table 3. We use a very short, 10-character shift for our sliding window. This is deliberate; it is meant to help us, effectively, "pan for gold"—to identify regions within specific books where there are high-probability stretches of memorized content. The approach that we take reveals the position (in character space) in each book where these regions occur (and if they occur at all). We explain the intention behind these experiments, and how the associated results should be interpreted.

Sequence length. We use 100-token sequences z, with a prefix (prompt) length of 50 tokens and a (target) suffix length of 50 tokens. We pick 100 tokens and a 50/50 split because this is accepted in the literature as a reasonable minimum for being confident that extracted suffixes are reflective of memorization (Carlini et al., 2023; 2021; Lee et al., 2022; Hayes et al., 2025b). 50 tokens for the prefix is the standard choice for 50-token suffixes, though is somewhat arbitrary. Of course, this prefix length uses an amount of context that is equivalent to the output, and so is compelling in this respect. However, when we are confident that our measurements are capturing valid instances of extraction, using longer prefixes can help *discover* more underlying memorization. (Indeed, this is part of where *discoverable* extraction gets its name; Carlini et al. (2023) demonstrate the "discoverability phenomenon" with respect to showing how extraction success can increase

with longer-context prompts.) We discuss this further (and include experiments with varied prefix lengths) in Section 4 and Appendix F. Other work uses shorter sequences (e.g., 64 tokens and a 32/32 token split; see Biderman et al. (2023)).

Sliding-window sequence chunking. For a given book, we start at the beginning of the text file in `Books3`. We sample a chunk of text that is sufficiently long to contain 100 tokens of corresponding tokenized text: to do this, we take a chunk of 800 characters, tokenize, take the first 100 tokens of the resulting tokenized sequence as the sequence, and discard the rest. (Each 100 token sequence is typically, but not always, reflective of 300-400 characters of text. Rare words and formatting make this vary considerably. This is why we use a character-chunk size of 800—to make sure we have sufficient head-room to always end up with a 100 token sequence.) We then shift 10 *characters* forward in the book text and repeat this process. We do this for the entire length of the book, which results in approximately one sequence every 10 characters, i.e., `len(book_characters)/10` total sequences z, for which we compute p_z. For example, *The Great Gatsby* has 270,870 characters, which results in roughly 27,000 sequences, for which we compute p_z (Equation 3). (See Section 2 and Appendix A.2.)

This means that the 100-token sequences overlap significantly. (Generally speaking, 100 tokens covers 300-400 characters, so shifting only 10 characters means there is a lot of overlap.) This is deliberate. We typically do not know how different open-weight models were trained; *a priori*, it is not clear exactly where we should begin sequences (or where they should end). With this approach, for extractable content, we expect to surface high-probability *regions* of memorized content, which we can then explore in more detail and more precisely in follow-up experiments. We picked a sliding window of 10 characters by running initial experiments with 1-, 10-, 50-, and 100-character windows. 10 characters exhibited a small loss in extraction signal (at $10\times$ cheaper cost than 1 character).

Decoding configuration Similar to Hayes et al. (2025b), after some initial tests with different decoding schemes, we use temperature $T = 1$ for top-k decoding with $k = 40$. Setting $T = 1$ makes the most sense for studying memorization, as this reflects the LLM's base probability distribution. Since extraction is only successful when the underlying training sequence's logits reflect very high probabilities (Section 2), it makes sense to clip the full-vocabulary logit distribution to reflect some number of the top-highest probability tokens per step. This is the same intuition behind using greedy decoding for discoverable extraction—but with a twist. Greedy decoding deterministically picks the single-most-likely token per step. (This is equivalent to top-k decoding with $k = 1$.) As the results in Hayes et al. (2025b) show, this can be a suboptimal decoding strategy for measuring extraction, for several reasons:

- Greedy decoding is rarely used with LLMs in practice. Non-deterministic sampling schemes are more reflective of realistic conditions. Because greedy decoding is deterministic, the same prompt always results in the same output. As a result, discoverable extraction provides (in this sense) 1-bit of information with respect to extraction: was it possible to extract the sequence? Using non-deterministic decoding schemes, we can get more detailed information about how extraction risk varies across sequences. (See Section 2 & Appendix A.)
- While greedy decoding picks the *locally* highest probability token per step, it is not guaranteed to result in the *globally* highest probability sequence (of the given length) under the model. Being able to explore more of the model's distribution with non-deterministic decoding schemes can surface sequences (of the same given length) that are higher probability. Such higher probability sequences can match the target suffix, where the greedy-decoded suffix does not (i.e., fails to extract that suffix).
- More generally, this allows for a bit more "wiggle room." By being able to pick beyond just the top-1 token per step, the decoding process take locally suboptimal steps, which can still result in high-probability sequences that (even if slightly lower probability than the greedy sequence, with respect to the normalized top-k distribution) reflect extraction. While we do not investigate this in detail in this work, we observe that (when this happens) the greedy-decoded suffix is a near-miss for the verbatim suffix (it is near-verbatim). It may be reasonable under more relaxed (near-verbatim) metrics for

extraction to have counted this greedy suffix as extracted, but that is not the approach we (or Hayes et al. (2025b)) take, and so these suffixes would be missed.

The above reasons also reflect the general motivations for using probabilistic extraction as our metric (Sections 2 & 7; Appendix A & Appendix B). Similar motivations for useful extraction measurements are also noted in earlier extraction papers (e.g., Carlini et al. (2021), Carlini et al. (2023)); however, this work typically opts to use greedy decoding based on the perception that it is cheaper to do so at scale. This work does not claim that greedy decoding is the best or only way to measure extraction, and in fact notes (Carlini et al., 2021)—and in select cases shows (Carlini et al., 2023)—that other (more expensive) decoding schemes are able to extract more training data.

Further, it turns out that it is actually *cheaper* to compute probabilistic extraction (with any top-k decoding configuration, including greedy/ $k = 1$), as we compute extraction probabilities as a post-processing operation directly off of the logits obtained from *inference*, rather than (as has been done in prior work (Carlini et al., 2021; 2023)) actually generating (greedy) completions. We can compute verbatim extraction probabilities with a single forward pass through the model, which has reduced overhead compared to greedy generation (Appendix A.3).

Note that because our implementation of probabilistic extraction makes computing log probabilities a post-processing operation on the output model logits (Appendix A), we also collect results for greedy-decoded discoverable extraction by setting $k = 1$, incurring negligible cost to do so. For brevity, we do not include detailed results for this setting. They under-count valid extraction, aligning with our findings in Section 3.2 and Appendix D, as well as Hayes et al. (2025b).

E.3 Book-specific results

For each of the 200 books we test, we include a separate subsection with the following information for 14 models (documented in Appendix E.1):

- **Heatmaps of extraction probabilities by location.** We include heatmaps (as in Figure 2b and Section 5). At each character, the heatmap plots the maximum extraction probability among overlapping 50-token suffixes spanning that character. These visualizations highlight "hot-spots" of high-probability extraction and show how such regions emerge and taper off within a book.
- **Histograms of extraction probabilities.** We plot the distribution over suffix extraction probabilities across all overlapping sequences from the sliding-window procedure (with a 10-character stride). These histograms illustrate how many sequences are extractable for a given model and how their extraction probabilities vary.

For histograms and heatmaps, results are presented in three groups corresponding to the categories of models described in Appendix E.1: models known with certainty to have been trained on `Books3` (blue), models that were likely trained on books also present in `Books3` (red), and one model (PHI 4) that, based on documentation, was not trained on whole copyrighted books (orange). See Section 4 and Appendix F for further discussion about how our results for all three categories reflect valid extraction.

We release an interactive, searchable version of our main results on 200 books and 14 models online, as well as a general-purpose coding suite for probabilistic extraction on GitHub.

E.3.1 *Things Fall Apart*, *Achebe*

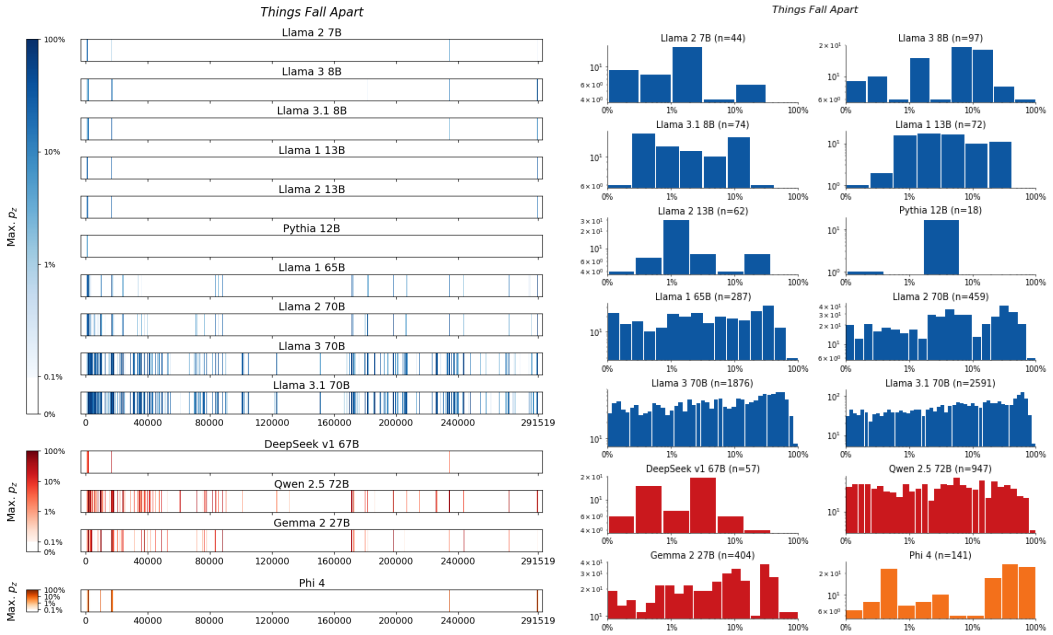


Figure 12: ***Things Fall Apart*, Achebe.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.2 *The Hitchhiker's Guide to the Galaxy - Omnibus*, *Adams*

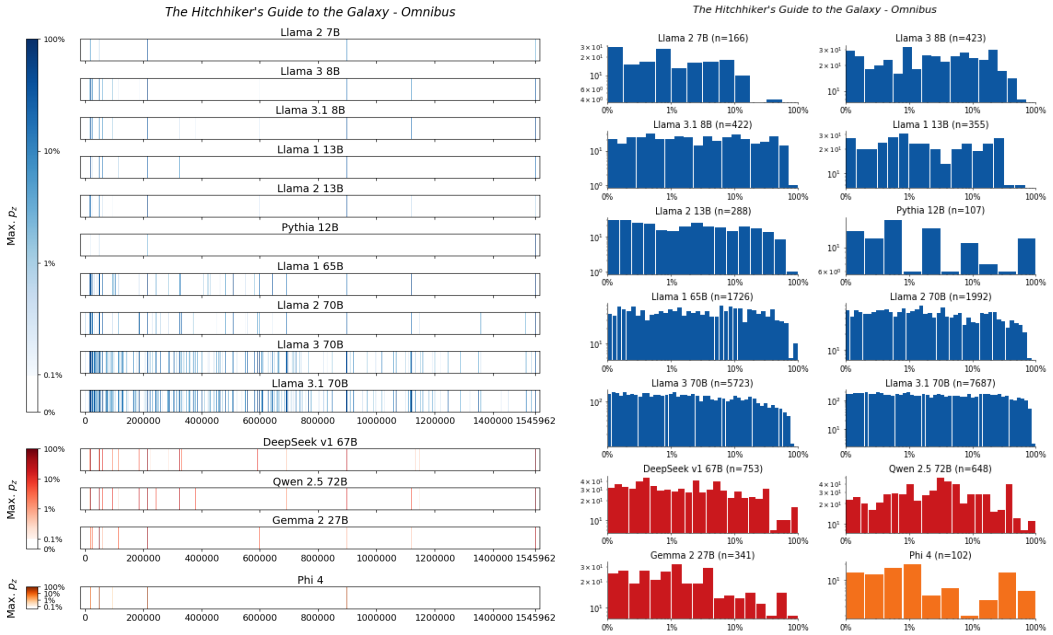


Figure 13: ***The Hitchhiker's Guide to the Galaxy - Omnibus*, Adams.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.3 *Americanah*, *Adichie*

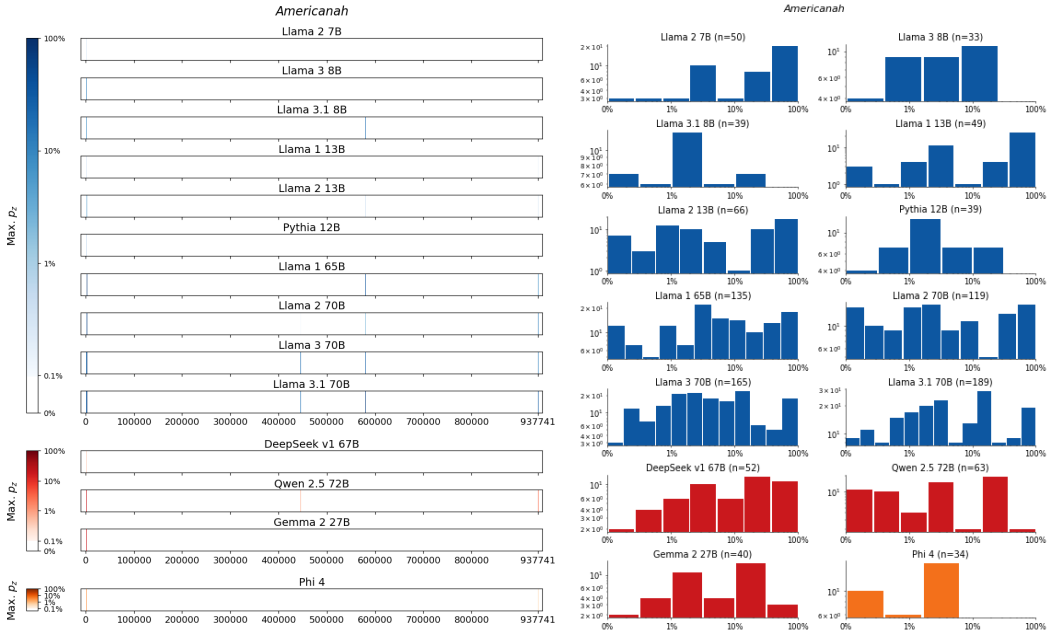


Figure 14: ***Americanah*, Adichie.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.4 *The Baghdad Clock*, *Al Rawi*

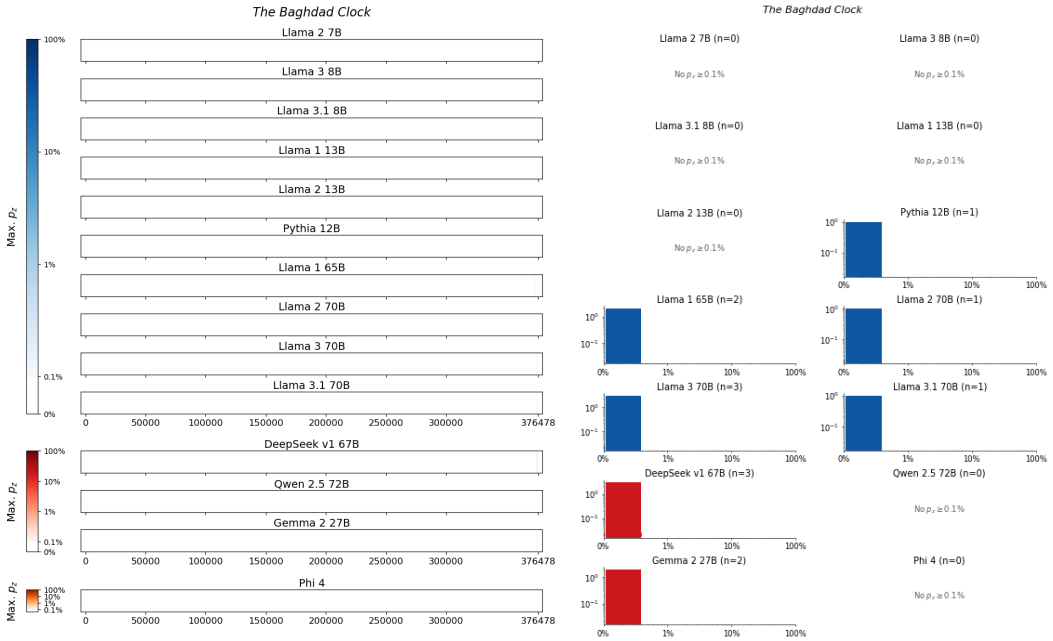


Figure 15: ***The Baghdad Clock*, Al Rawi.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.5 *Industrial Magic, Armstrong*

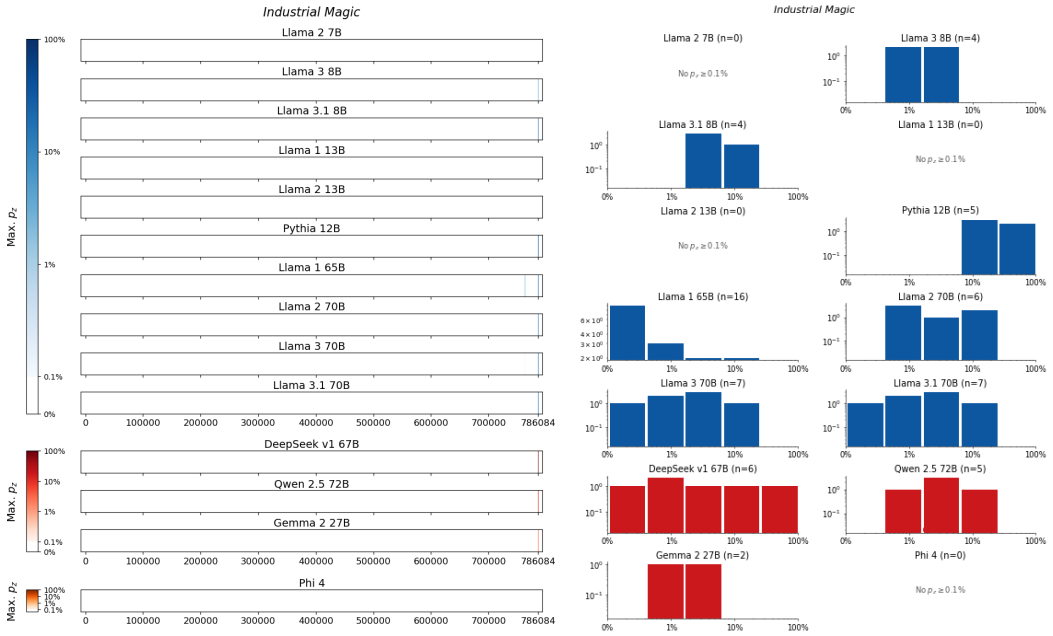


Figure 16: ***Industrial Magic, Armstrong.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.6 *Fantastic Voyage, Asimov*

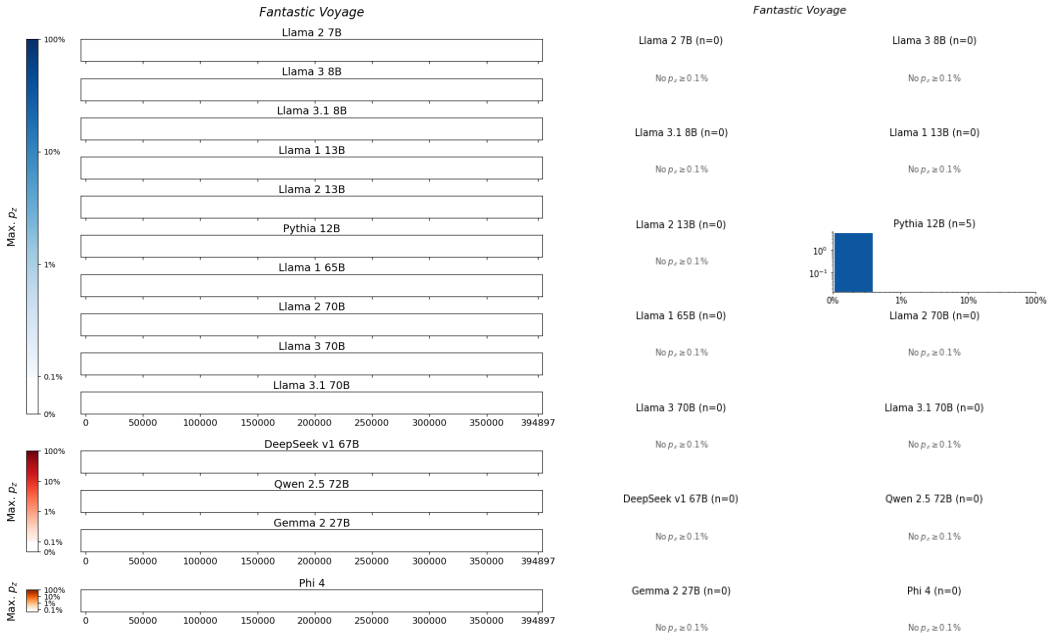


Figure 17: ***Fantastic Voyage, Asimov.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.7 *The Complete Robot, Asimov*

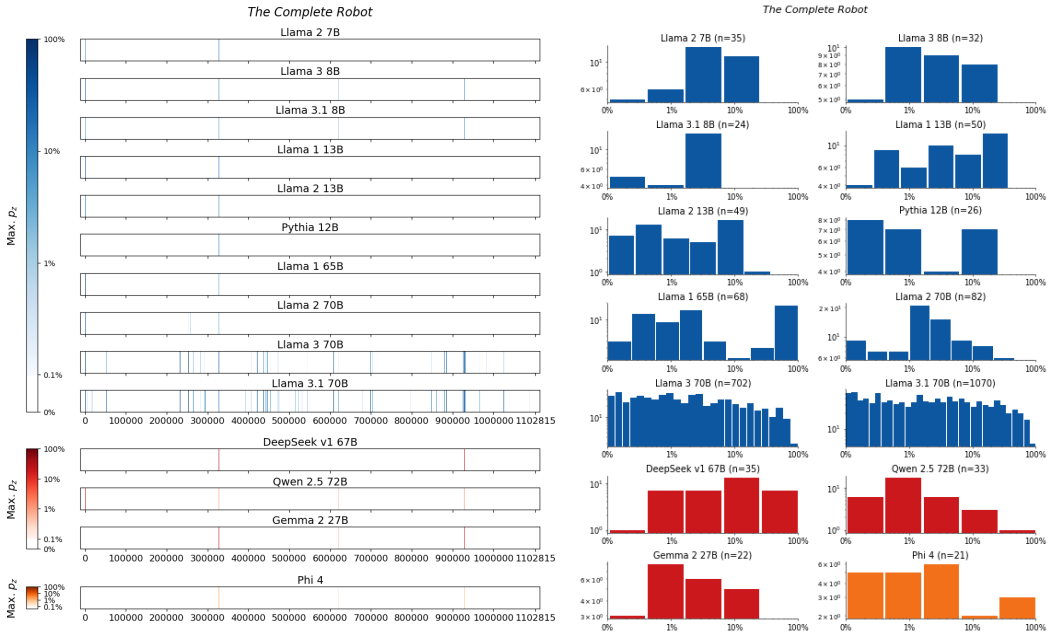


Figure 18: ***The Complete Robot*, Asimov.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.8 *The Handmaid's Tale, Atwood*

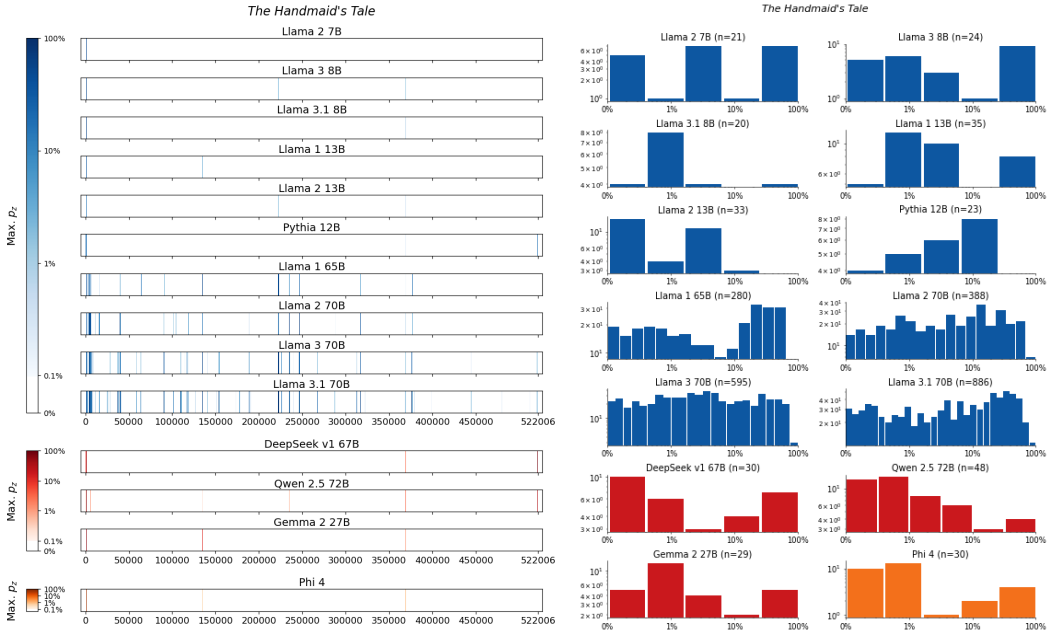


Figure 19: ***The Handmaid's Tale*, Atwood.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.9 *Pride and Prejudice*, *Austen*

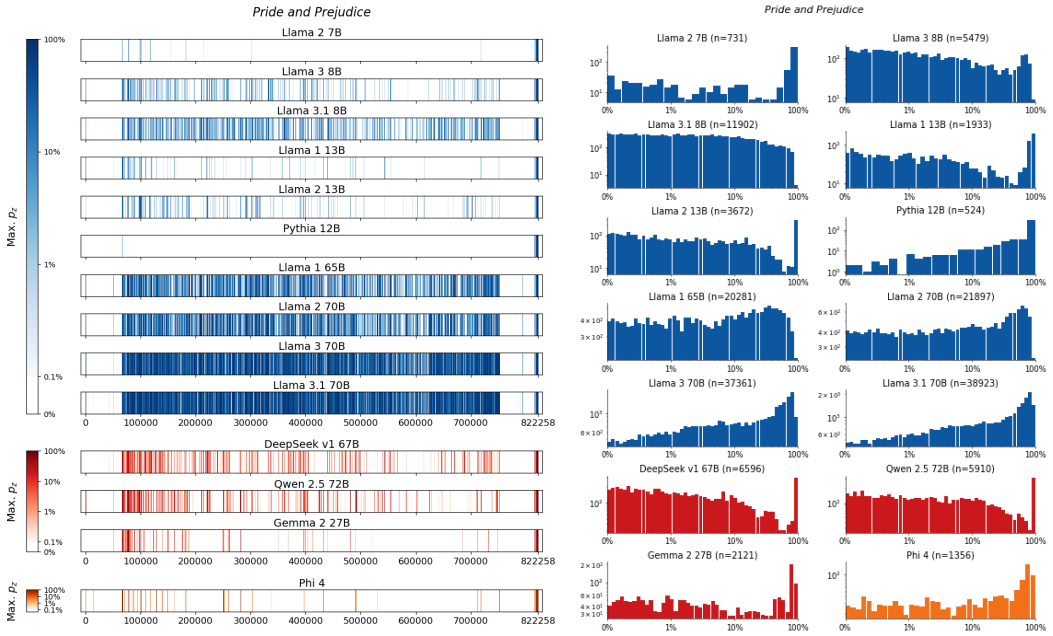


Figure 20: ***Pride and Prejudice*, Austen.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$). This is a public domain book, so PHI 4 is not a negative control.

E.3.10 *The Christmas Train*, *Baldacci*

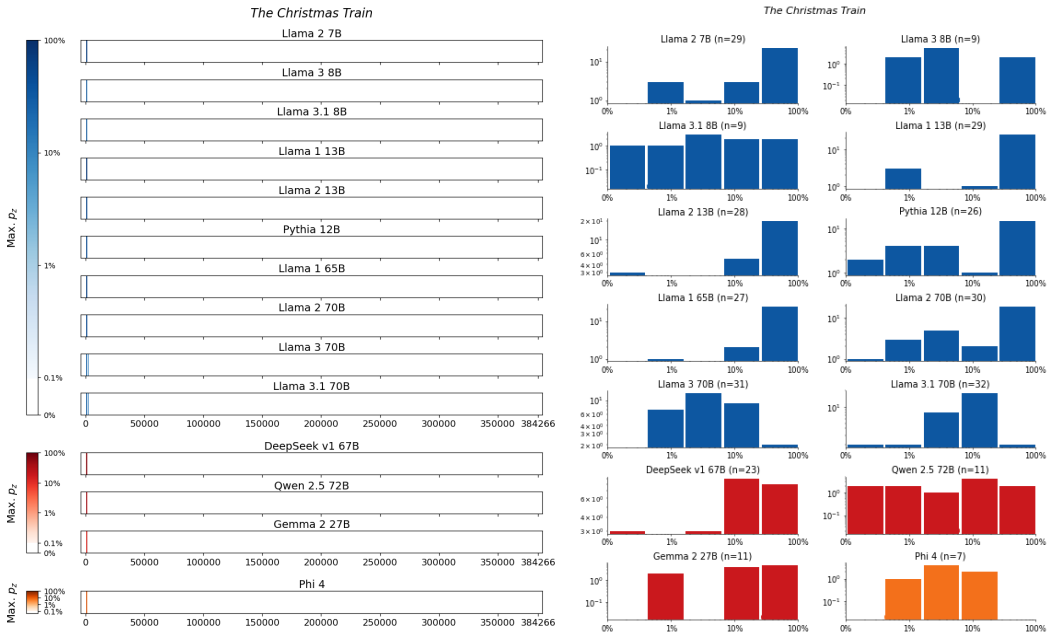


Figure 21: ***The Christmas Train*, Baldacci.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.11 *Notes of a Native Son*, *Baldwin*

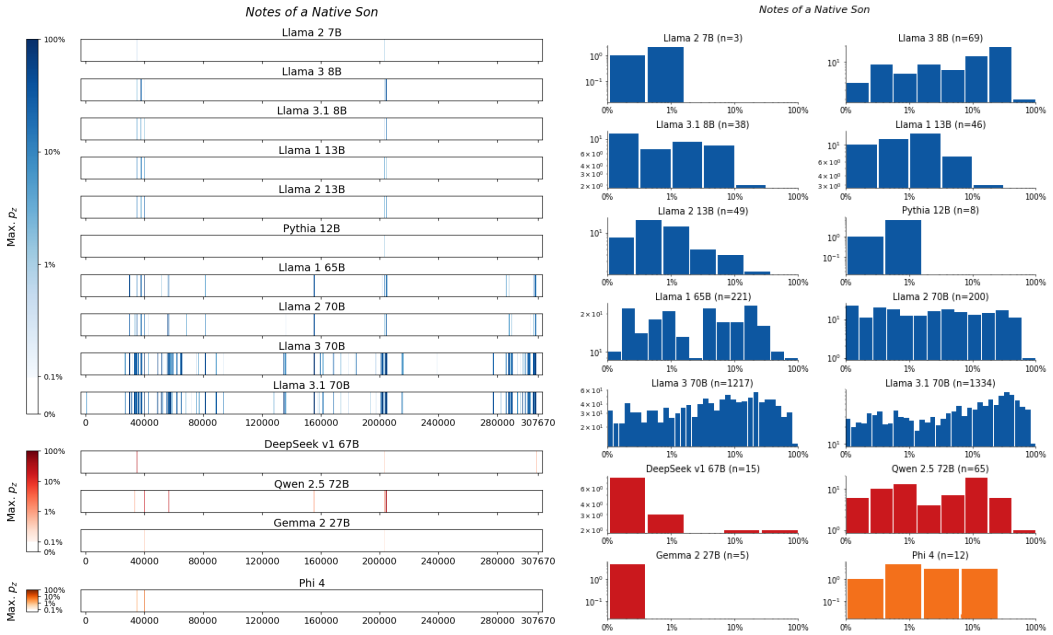


Figure 22: ***Notes of a Native Son*, Baldwin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.12 *Another Country*, *Baldwin*

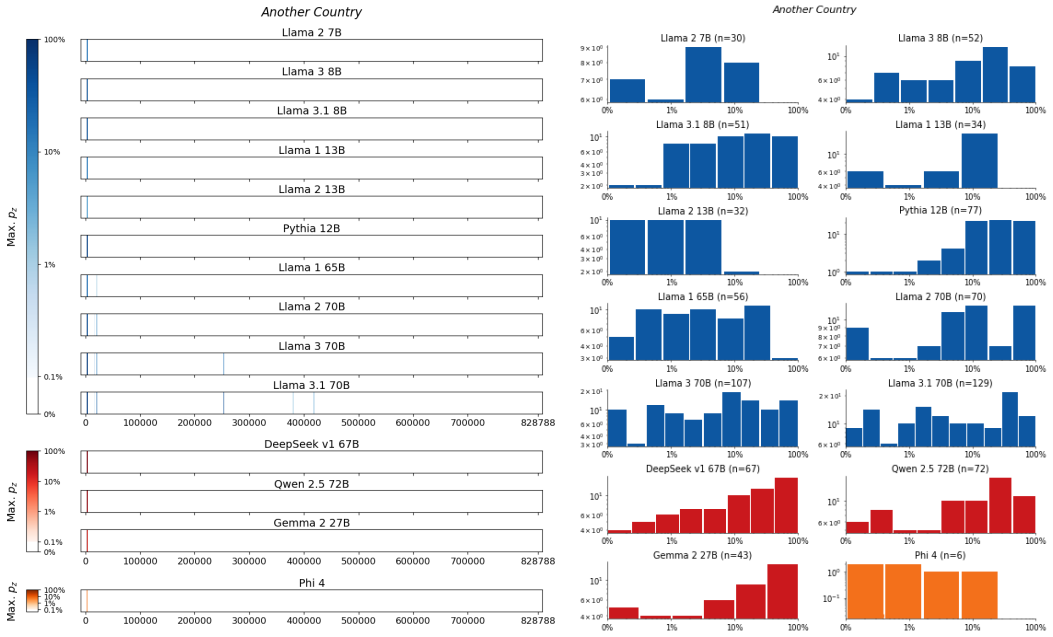


Figure 23: ***Another Country*, Baldwin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.13 *The Lemon Table, Barnes*

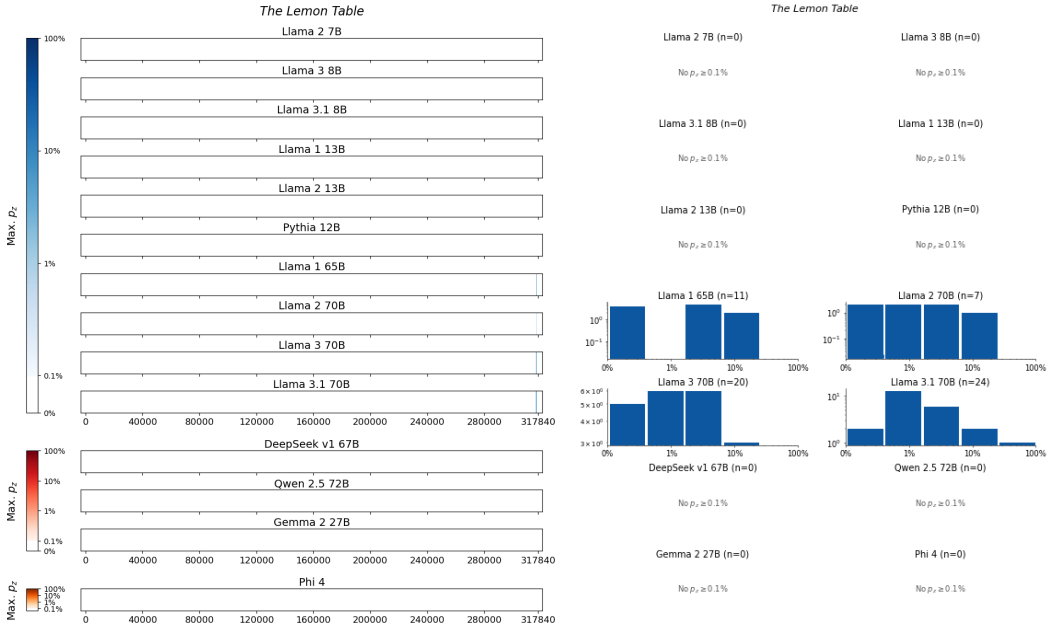


Figure 24: ***The Lemon Table*, Barnes.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.14 *Dante and the Origins of Italian Literary Culture, Barolini*

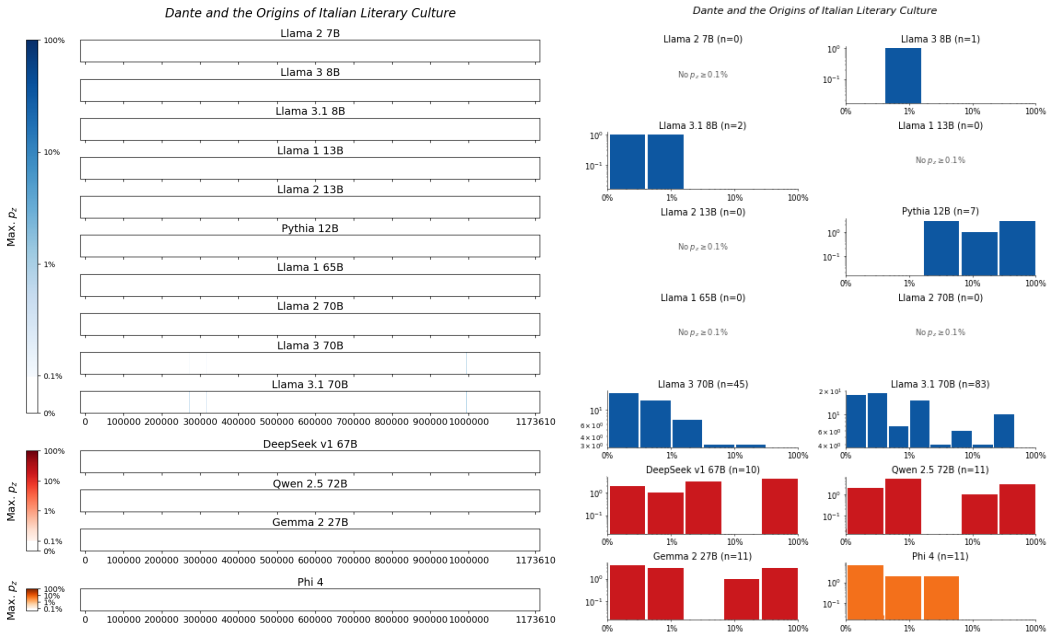


Figure 25: ***Dante and the Origins of Italian Literary Culture*, Barolini.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.15 The Parthenon, Beard

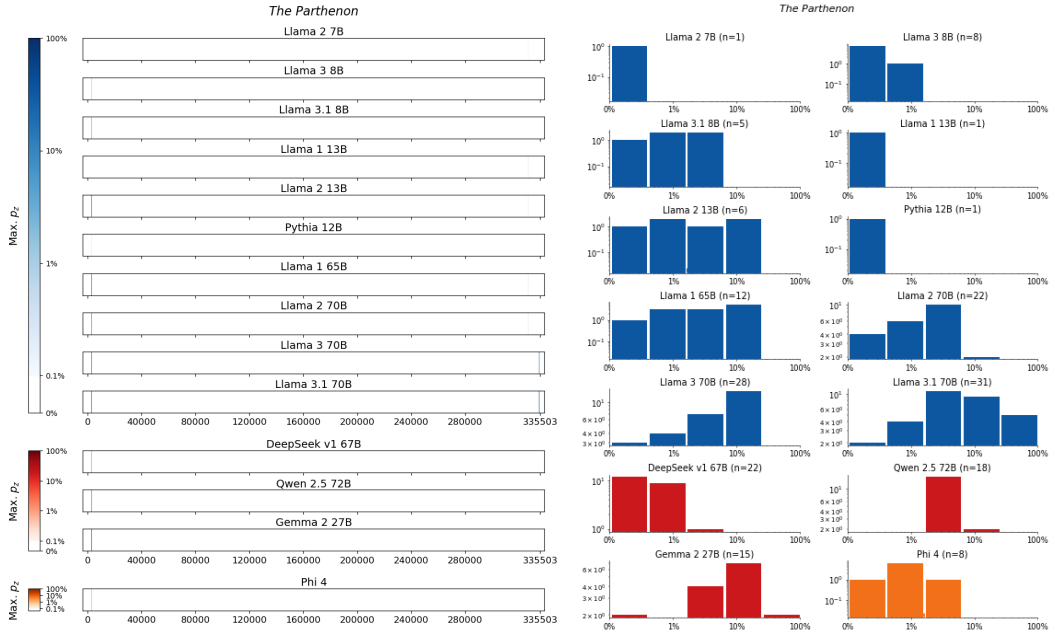


Figure 26: ***The Parthenon*, Beard.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.16 Guam: Past and Present, Beardsley

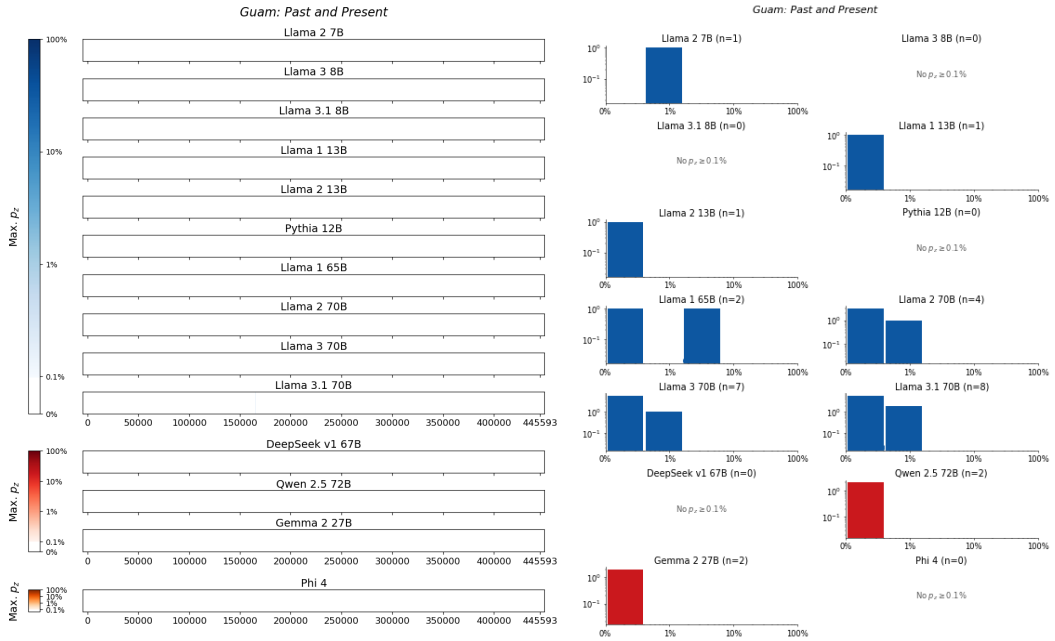


Figure 27: ***Guam: Past and Present*, Beardsley.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.17 Waiting for Godot, Beckett

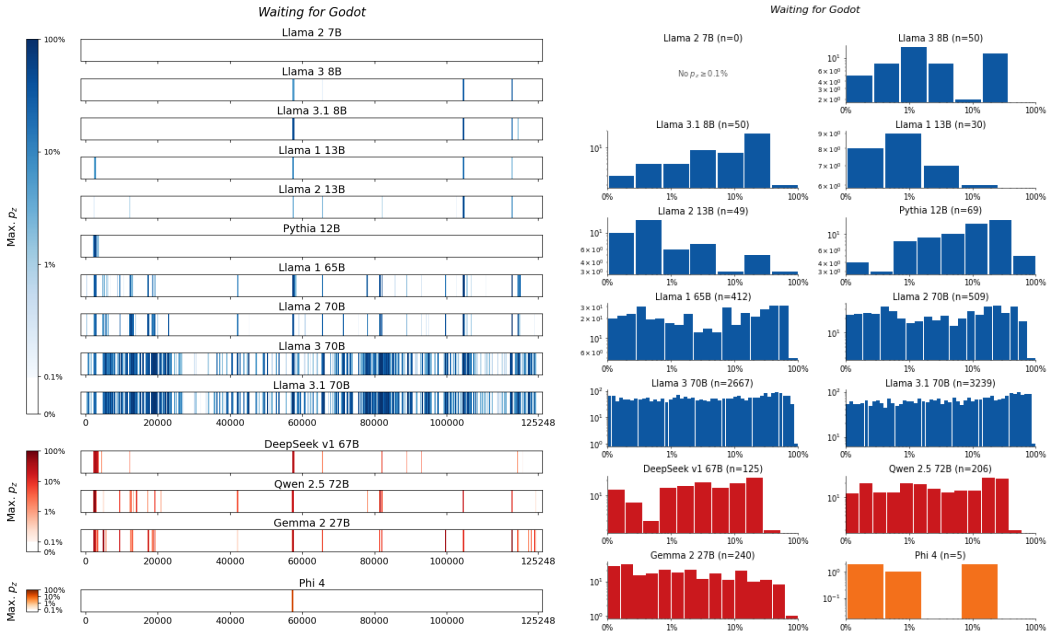


Figure 28: ***Waiting for Godot*, Beckett.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.18 The Lonely Soldier, Benedict

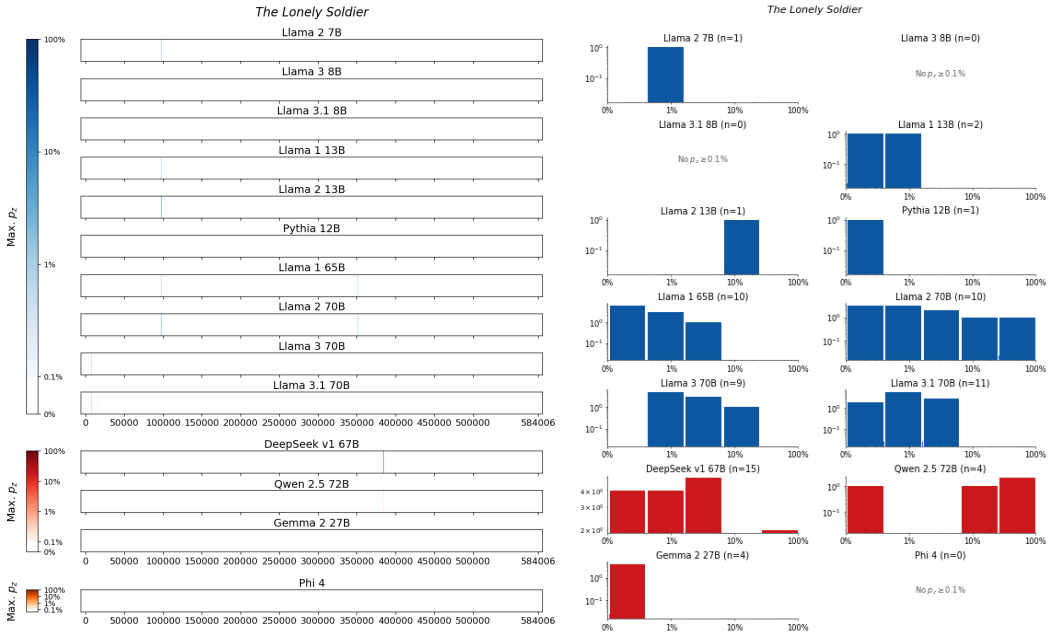


Figure 29: ***The Lonely Soldier*, Benedict.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.19 *Simple Cakes, Berry*

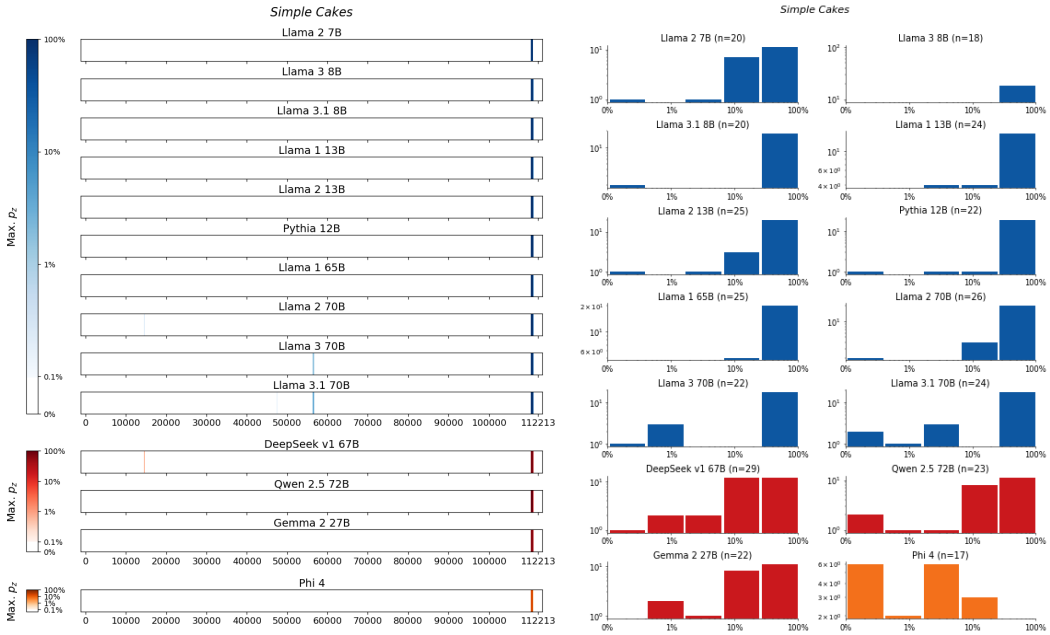


Figure 30: ***Simple Cakes, Berry.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.20 *Paradise Valley, Box*

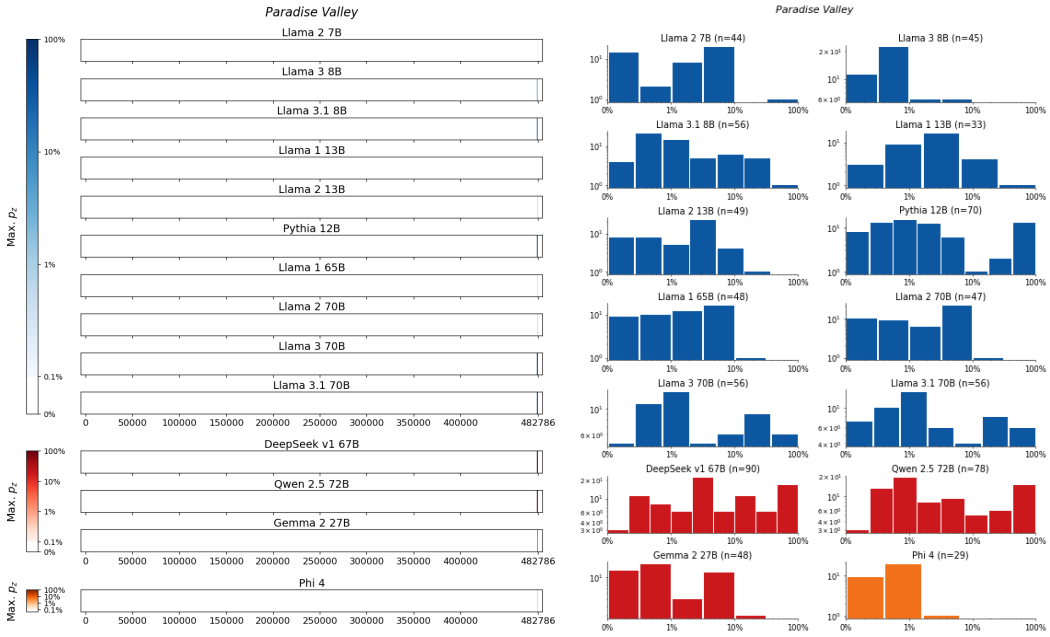


Figure 31: ***Paradise Valley, Box.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.21 The Cat's Pajamas, Bradbury

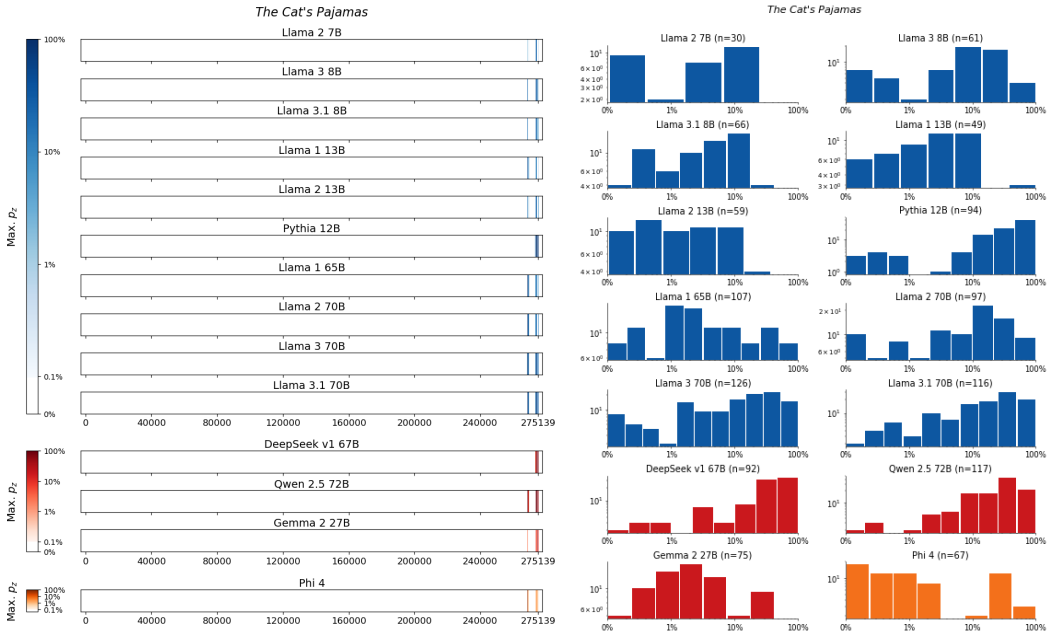


Figure 32: ***The Cat's Pajamas*, Bradbury.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.22 London in Chains, Bradshaw

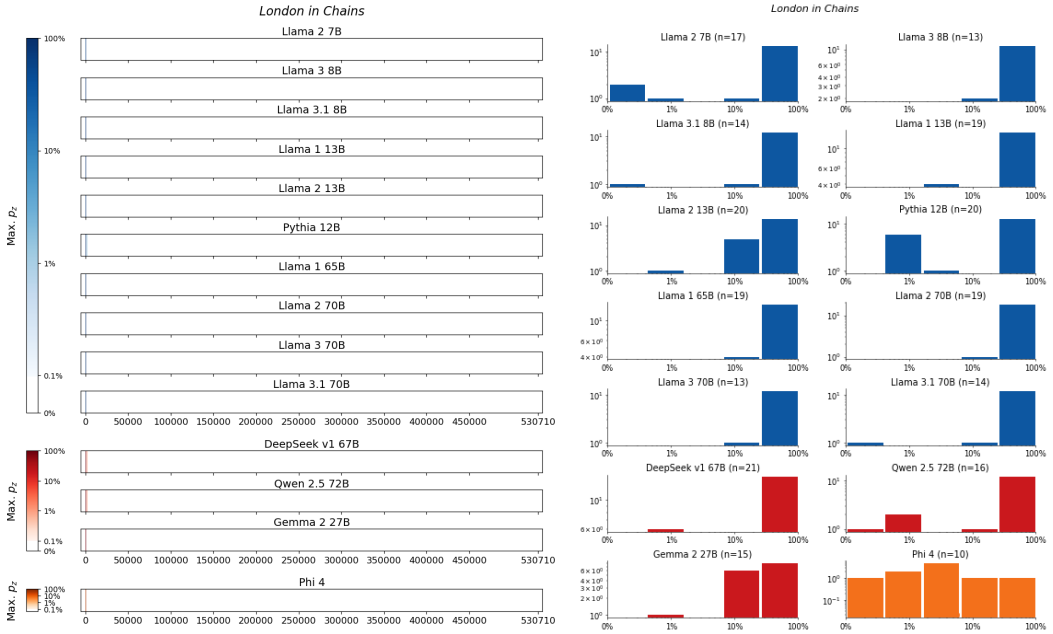


Figure 33: ***London in Chains*, Bradshaw.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.23 *My Einstein,* Brockman

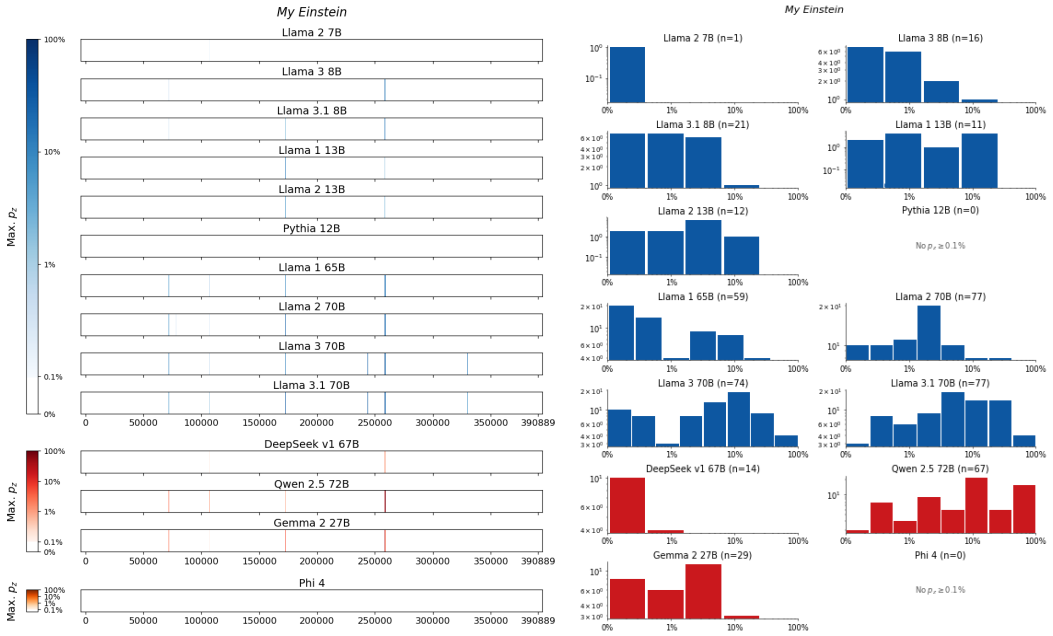


Figure 34: ***My Einstein*, Brockman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.24 *The Da Vinci Code,* Brown

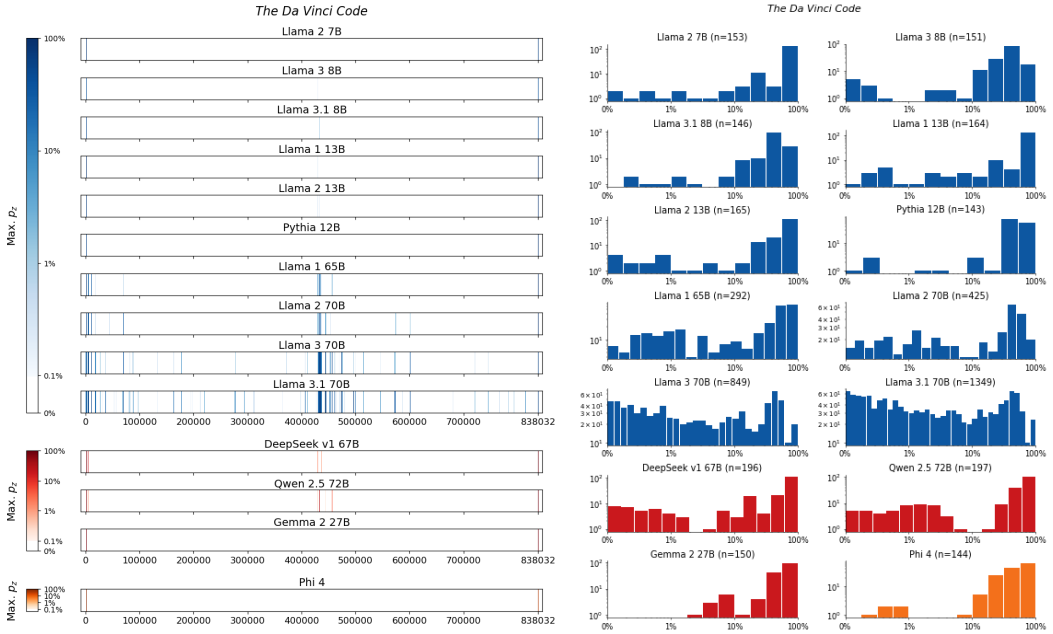


Figure 35: ***The Da Vinci Code*, Brown.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.25 *Live and Learn, Bryant*

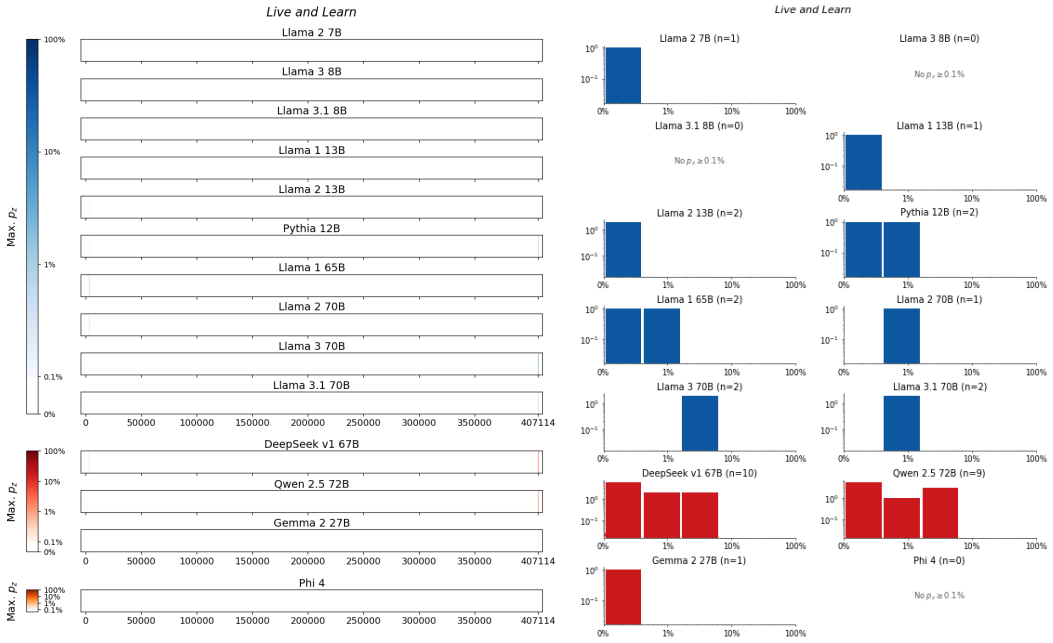


Figure 36: ***Live and Learn*, Bryant.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.26 *Knowing Your Value, Brzezinski*

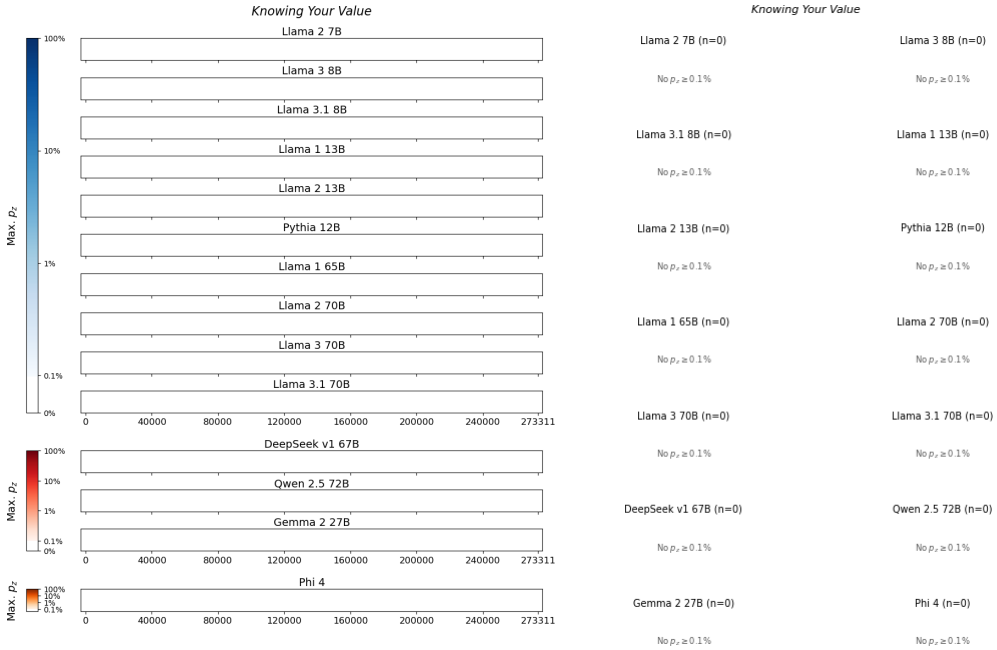


Figure 37: ***Knowing Your Value*, Brzezinski.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.27 *The Myth of Sisyphus*, *Camus*



Figure 38: ***The Myth of Sisyphus*, Camus.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.28 *Alice's Adventures in Wonderland*, *Carroll*

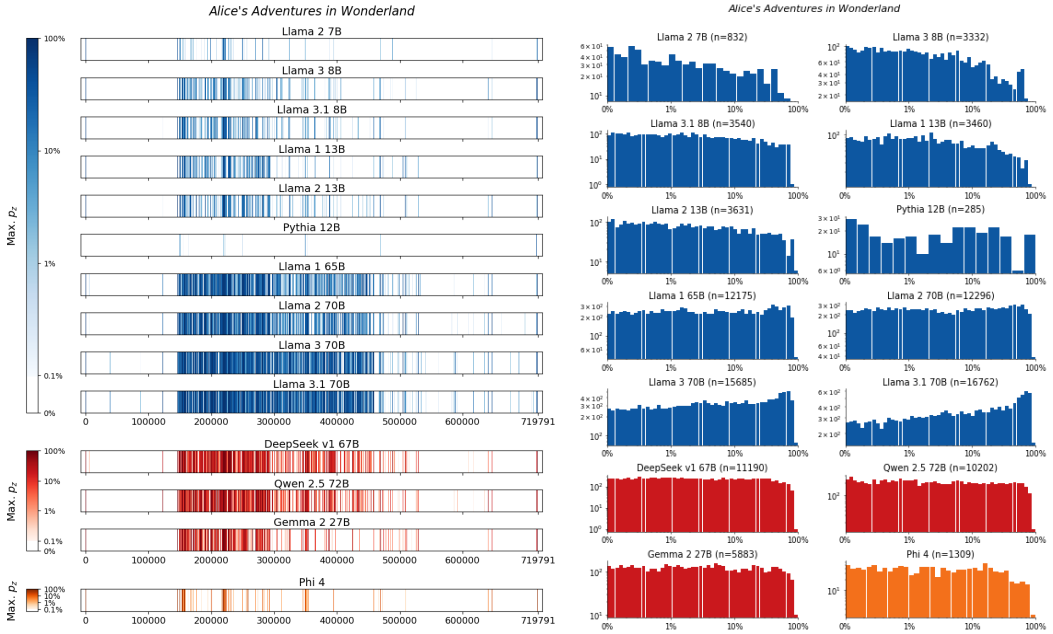


Figure 39: ***Alice's Adventures in Wonderland*, Carroll.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$). This is a public domain book, so PHI 4 is not a negative control.

E.3.29 *The Infinity Link, Carver*

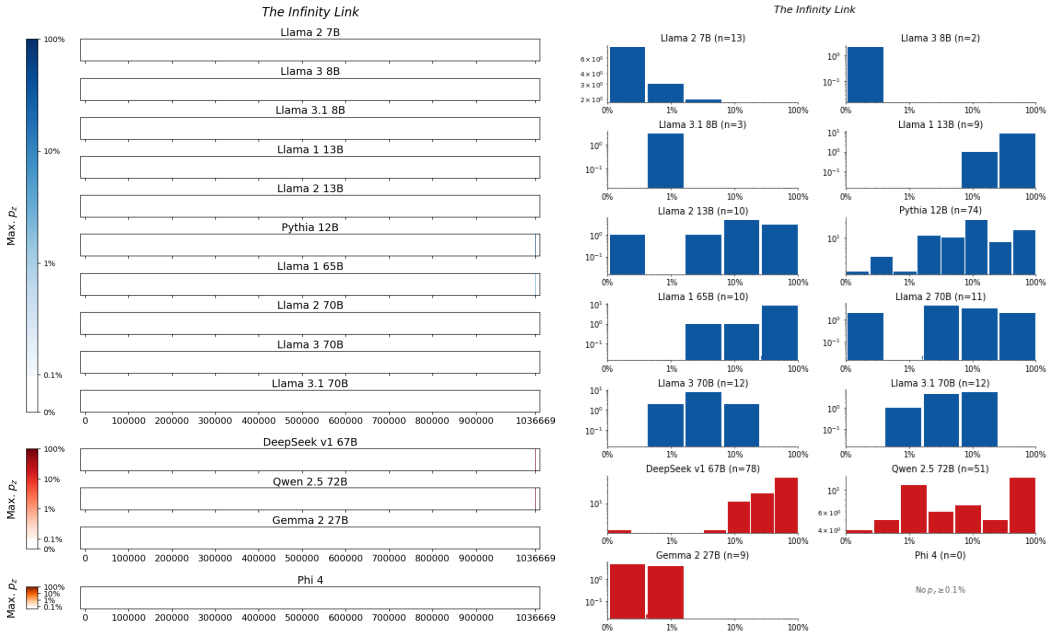


Figure 40: ***The Infinity Link, Carver.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.30 *Murder on the Orient Express, Christie*

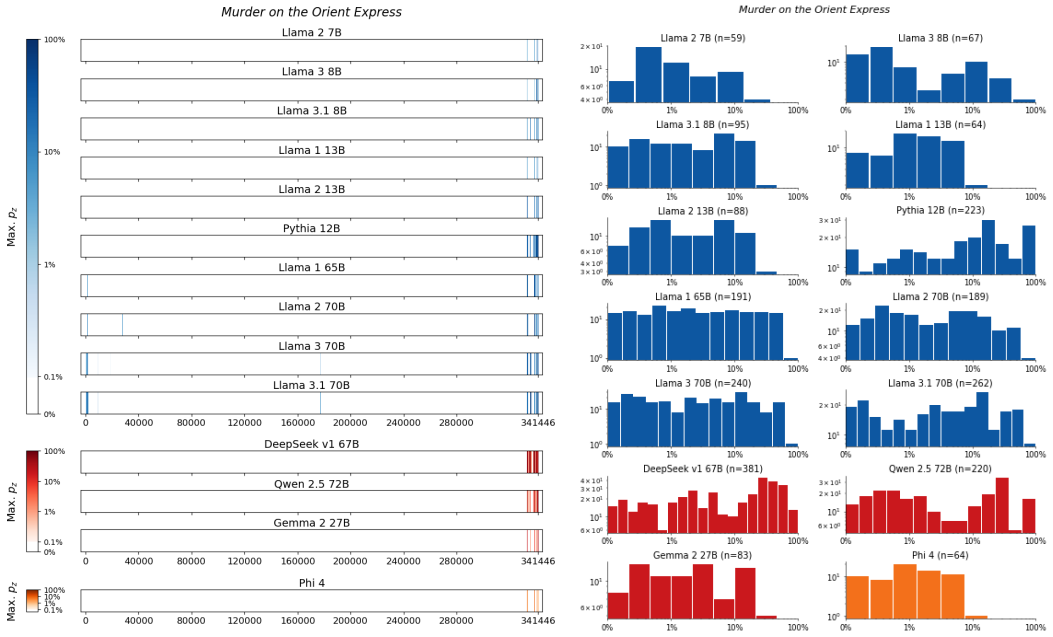


Figure 41: ***Murder on the Orient Express, Christie.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.31 *And Then There Were None*, *Christie*

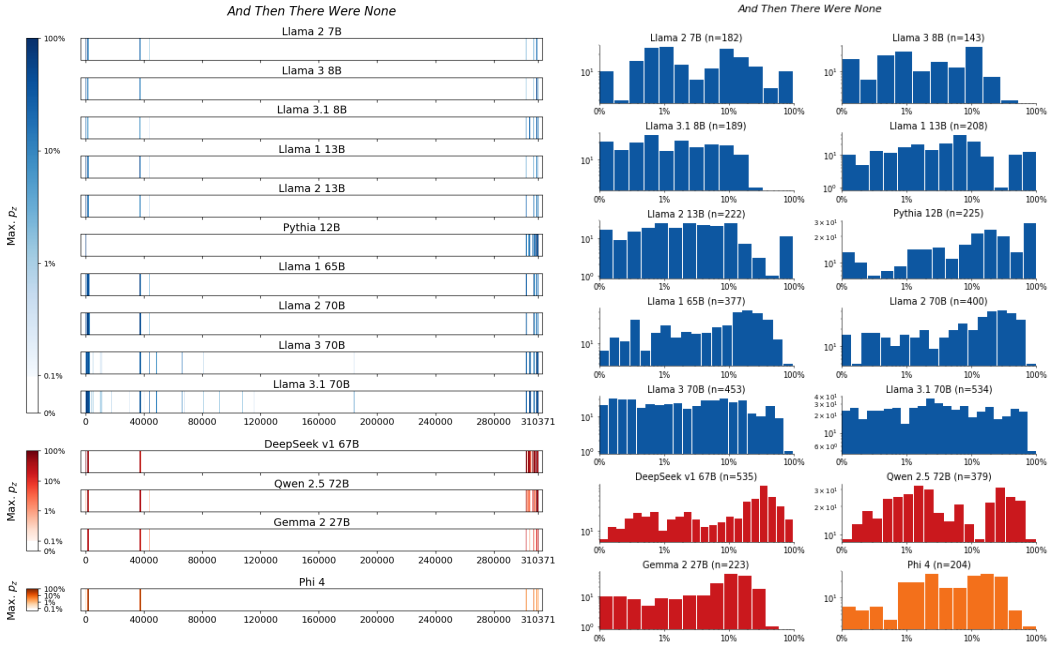


Figure 42: ***And Then There Were None*, Christie.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.32 *The Beautiful Struggle*, *Coates*

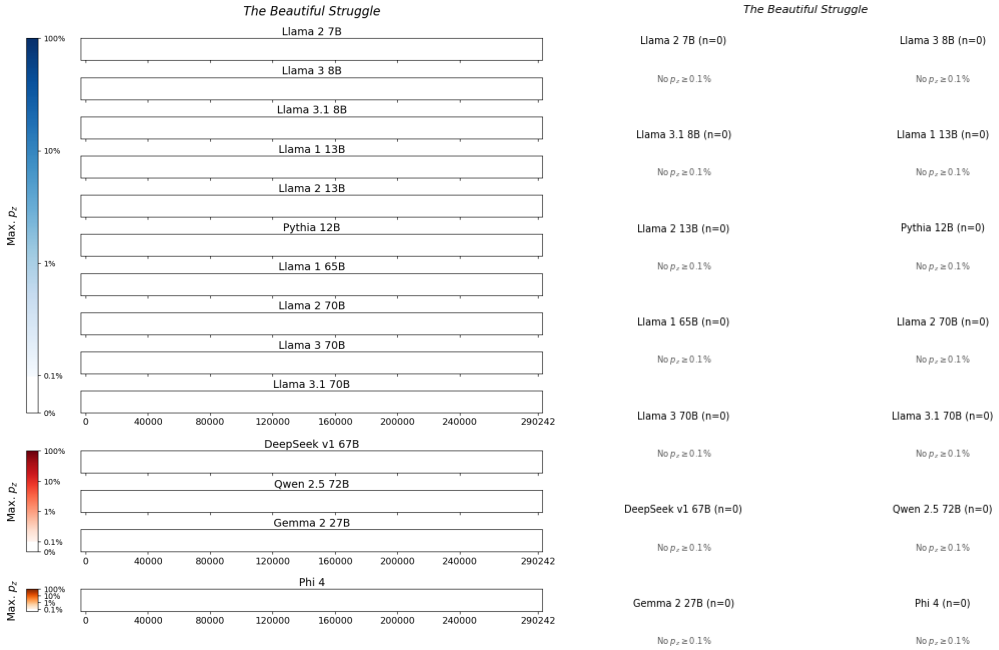


Figure 43: ***The Beautiful Struggle*, Coates.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.33 *We Were Eight Years in Power,* Coates

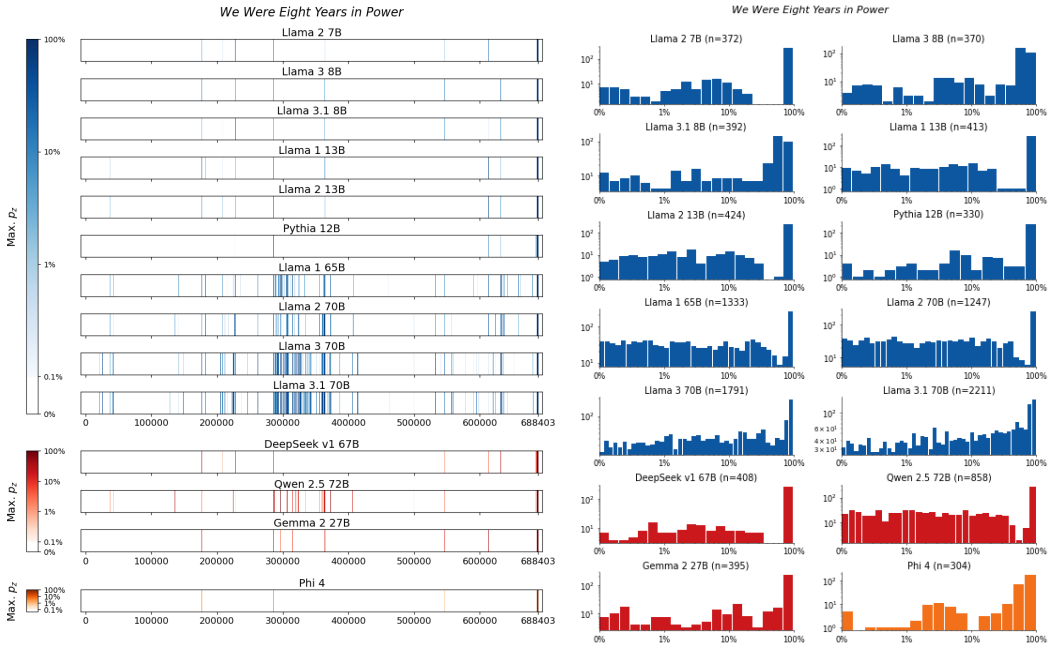


Figure 44: ***We Were Eight Years in Power,* Coates.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.34 *The Water Dancer,* Coates

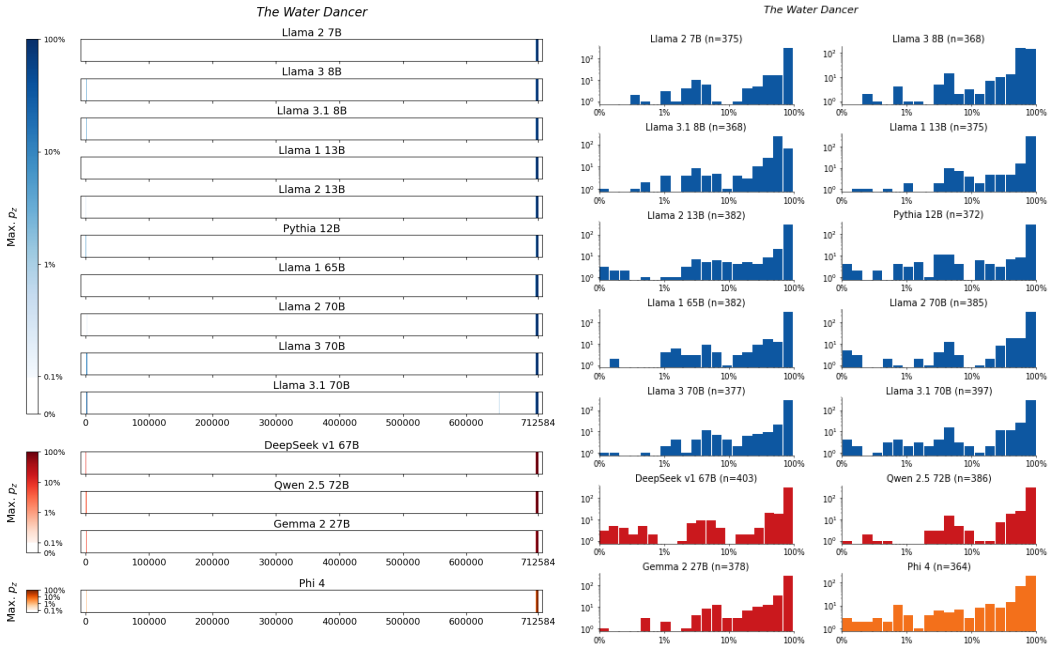


Figure 45: ***The Water Dancer,* Coates.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.35 *The Infernal Machine*, *Cocteau*

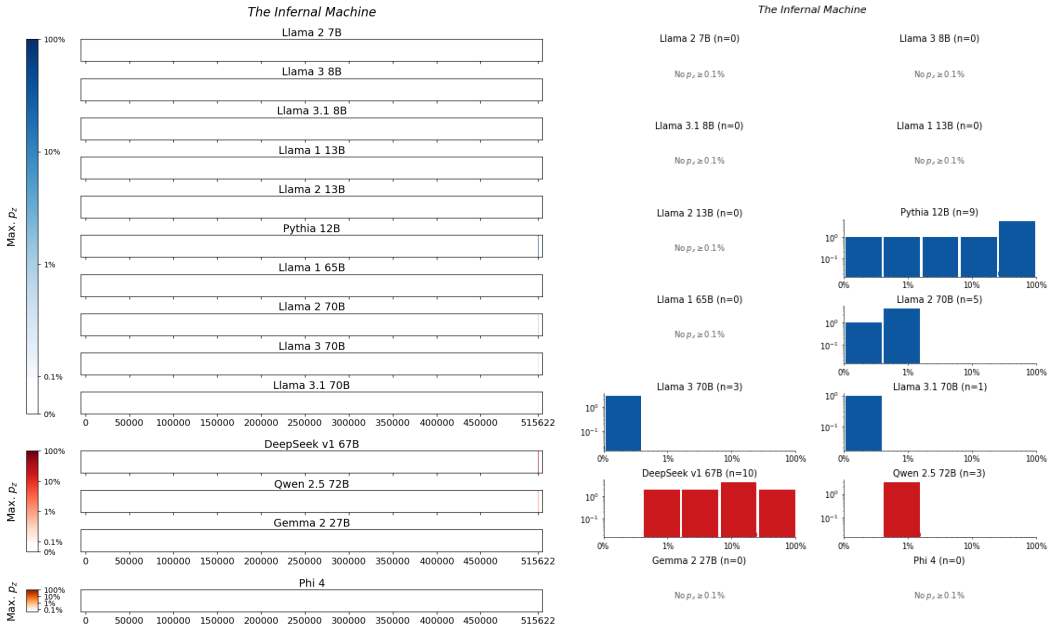


Figure 46: ***The Infernal Machine*, Cocteau.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.36 *The Alchemist*, *Coelho*

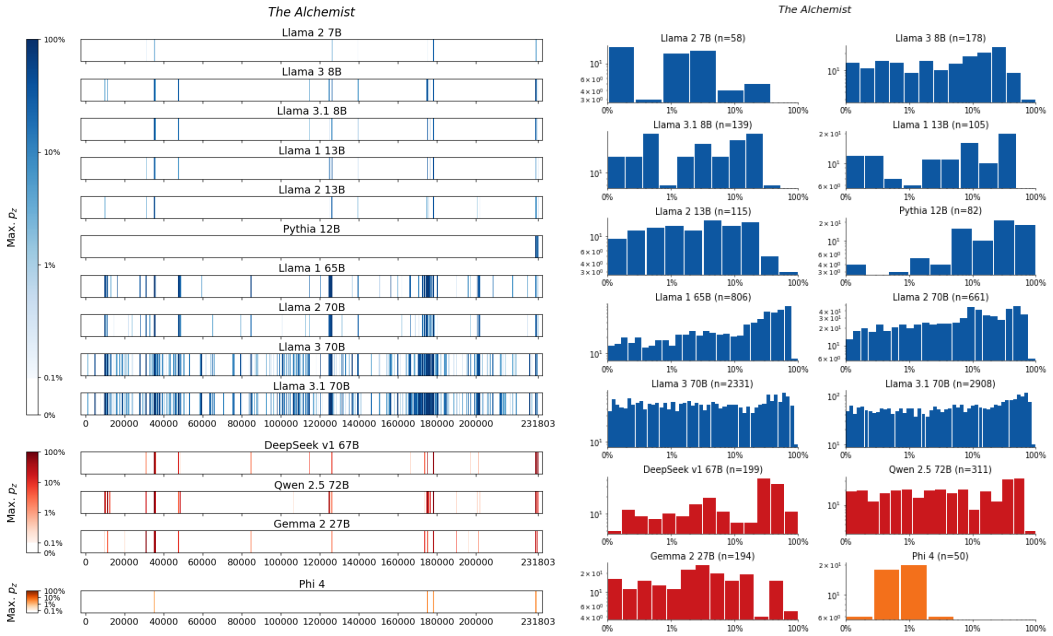


Figure 47: ***The Alchemist*, Coelho.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.37 Dungeons and Dragons and Philosophy, Cogburn & Silcox

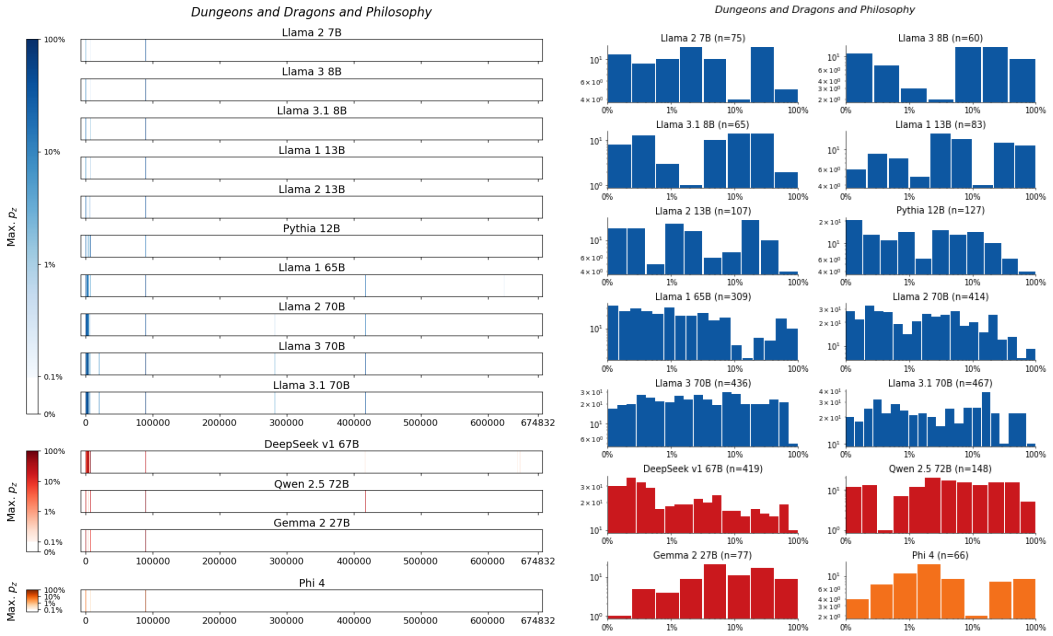


Figure 48: ***Dungeons and Dragons and Philosophy*, Cogburn & Silcox.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.38 Mark Rothko, Cohen-Solal

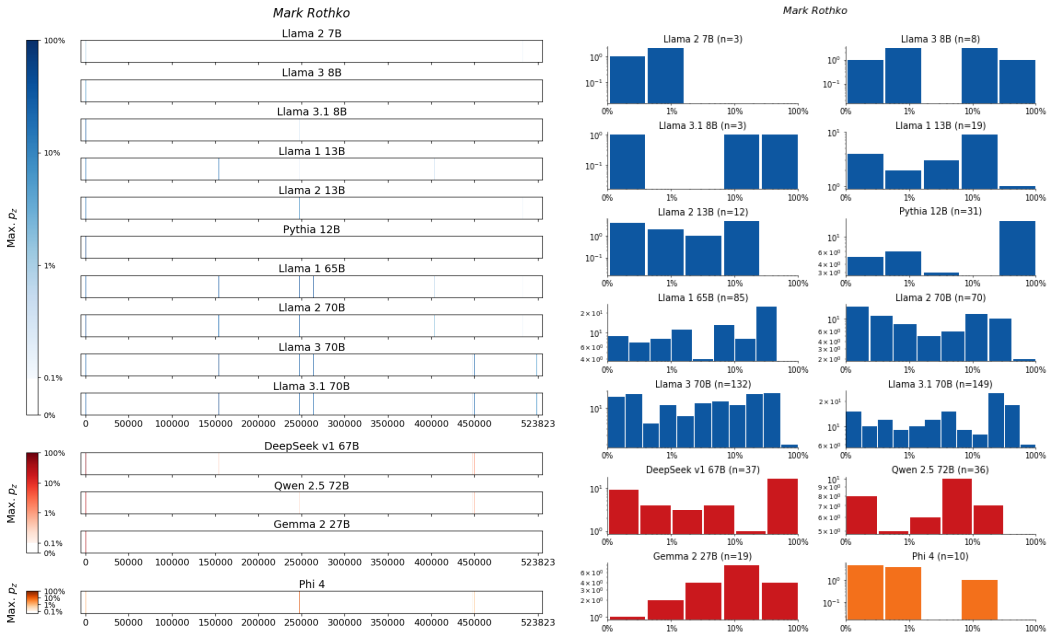


Figure 49: ***Mark Rothko*, Cohen-Solal.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.39 *The Hunger Games*, *Collins*

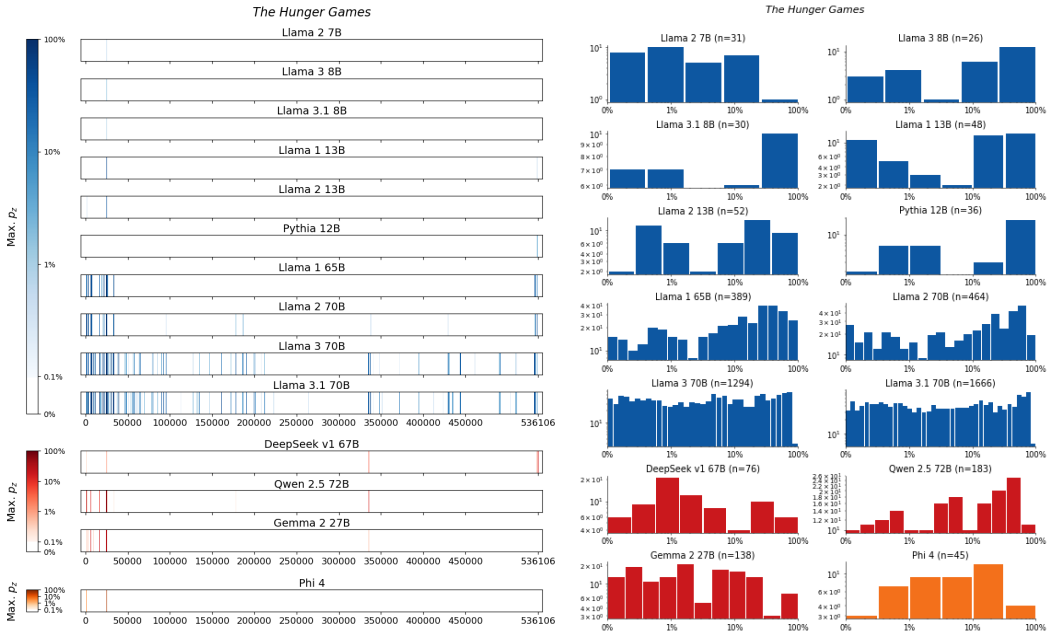


Figure 50: ***The Hunger Games*, Collins.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.40 *The Dragon Never Sleeps*, *Cook*

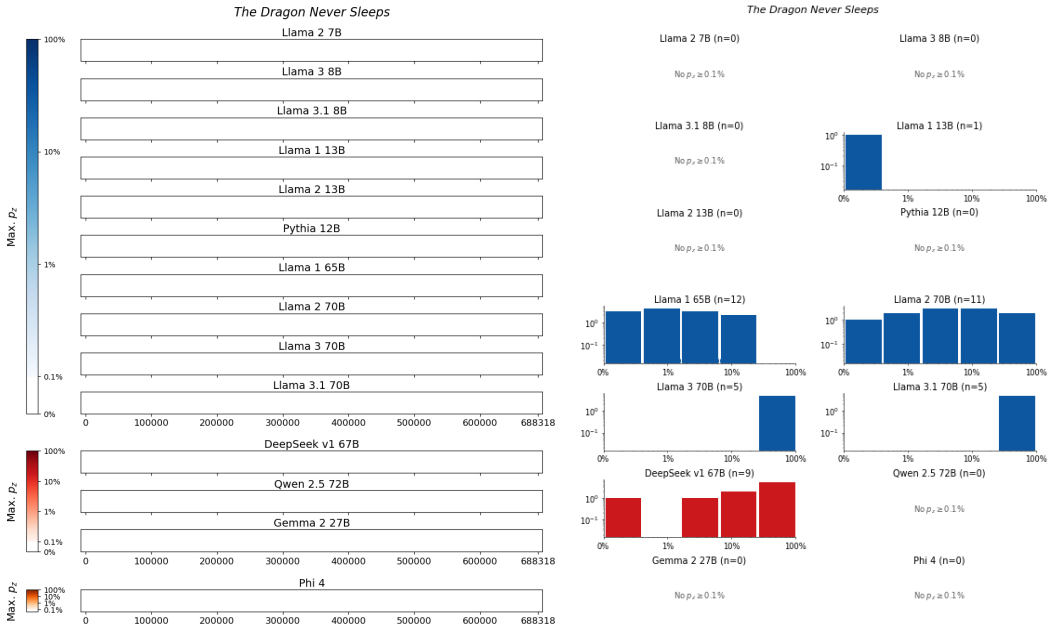


Figure 51: ***The Dragon Never Sleeps*, Cook.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.41 *The 7 Habits of Highly Effective People, Covey*

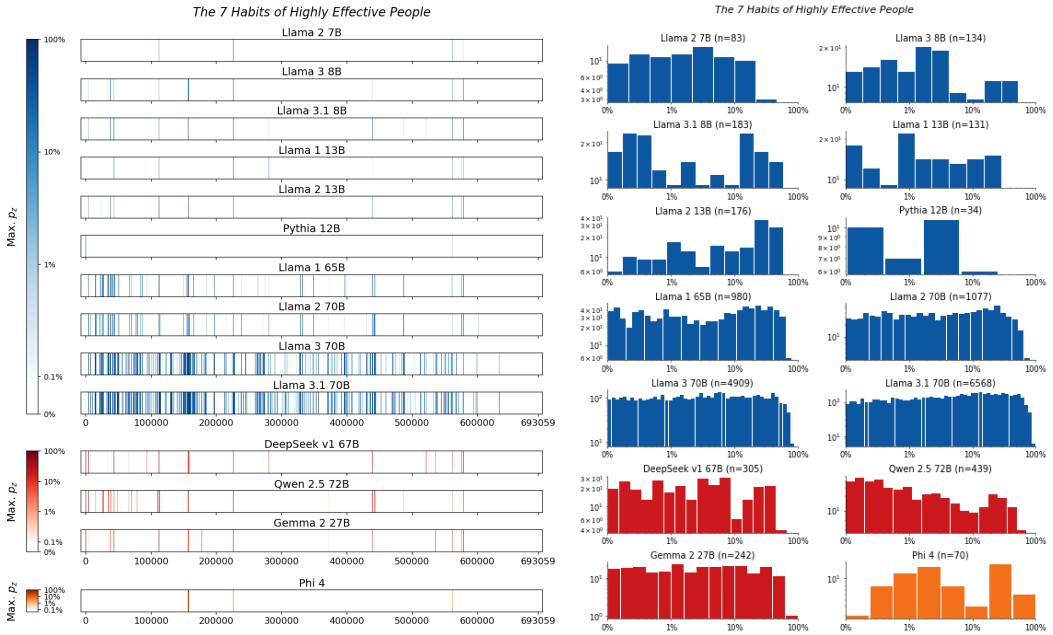


Figure 52: ***The 7 Habits of Highly Effective People*, Covey.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.42 *Bad Kid, Crabb*

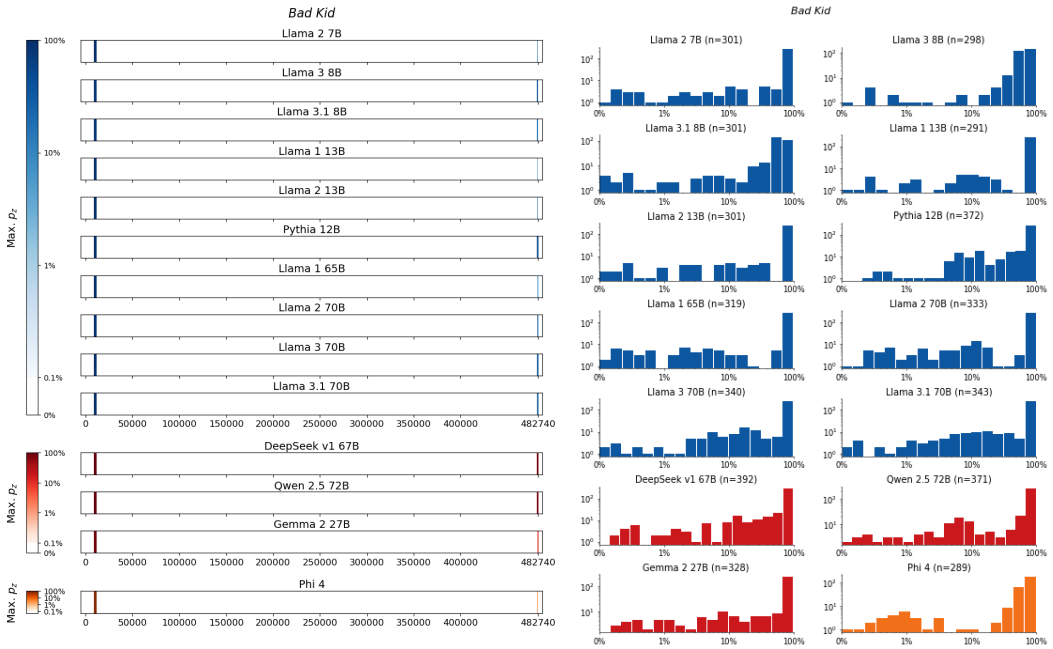


Figure 53: ***Bad Kid*, Crabb.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.43 *Lullaby Town, Crais*

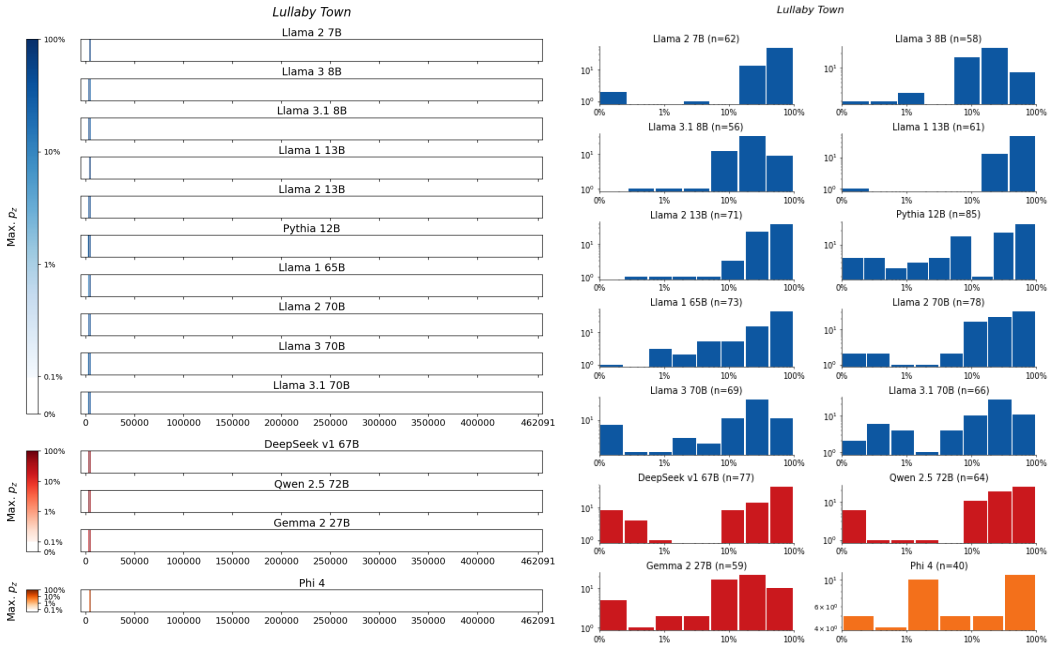


Figure 54: ***Lullaby Town, Crais.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.44 *Jurassic Park, Crichton*

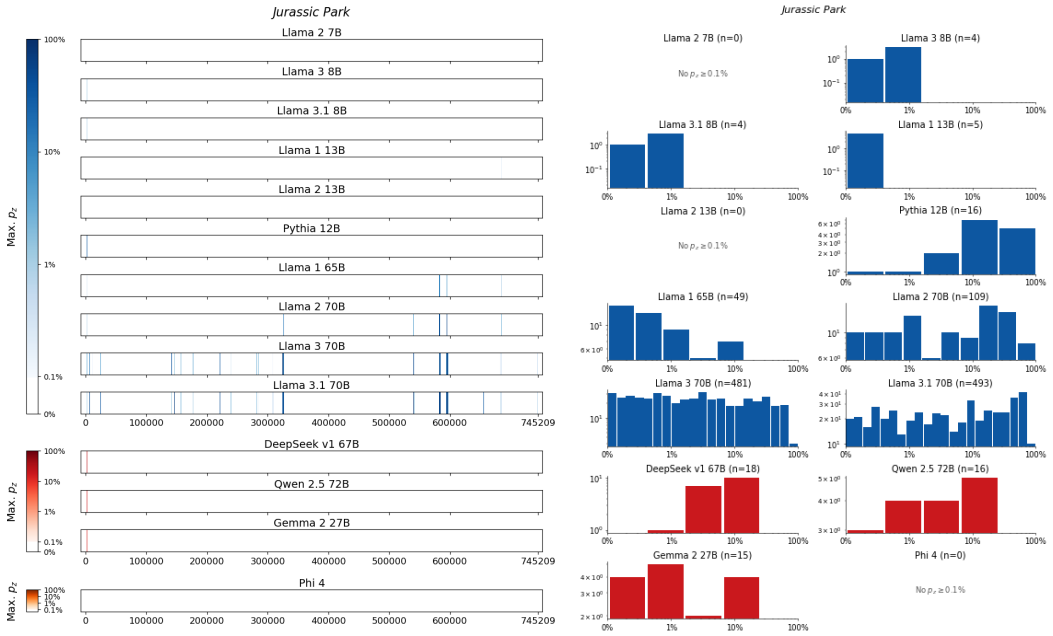


Figure 55: ***Jurassic Park, Crichton.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.45 *The Hours, Cunningham*

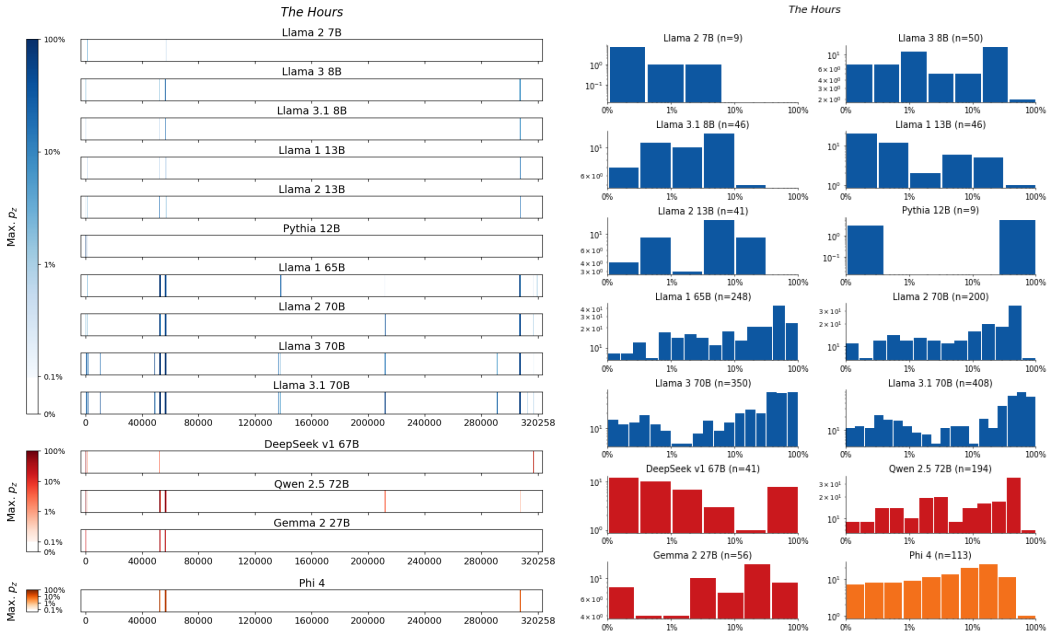


Figure 56: ***The Hours*, Cunningham.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.46 *Inhuman Land, Czapski*

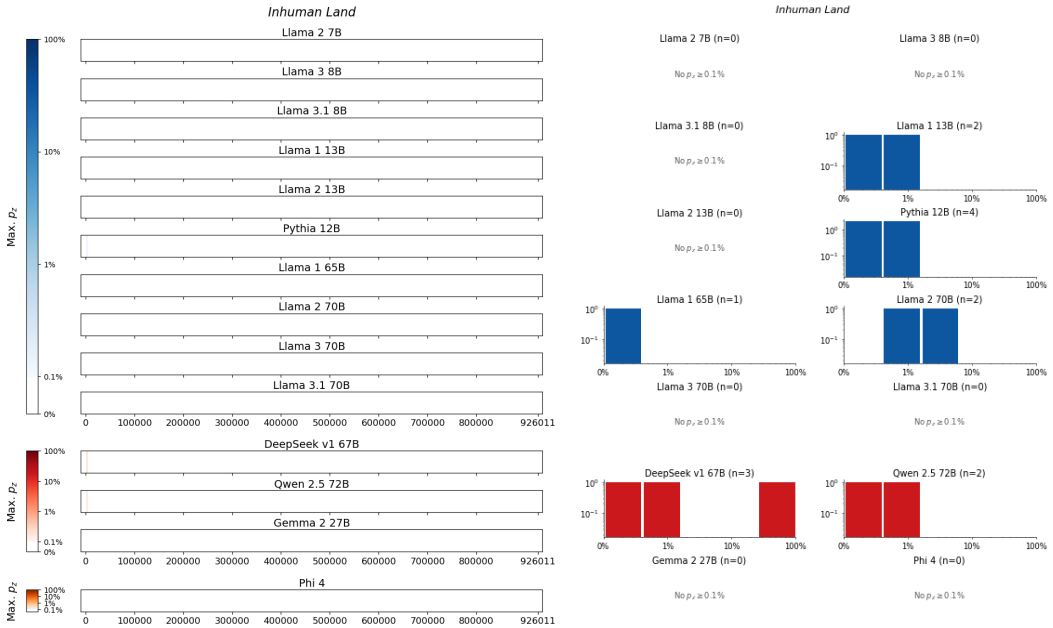


Figure 57: ***Inhuman Land*, Czapski.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.47 *Charlie and the Chocolate Factory,* Dahl

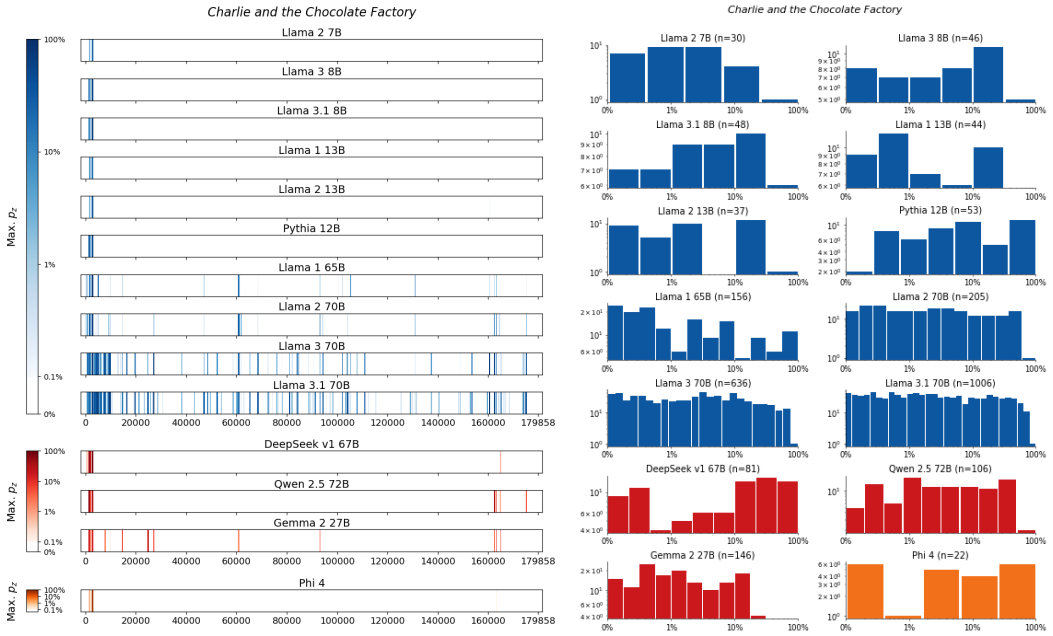


Figure 58: ***Charlie and the Chocolate Factory,* Dahl.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.48 *James and the Giant Peach,* Dahl

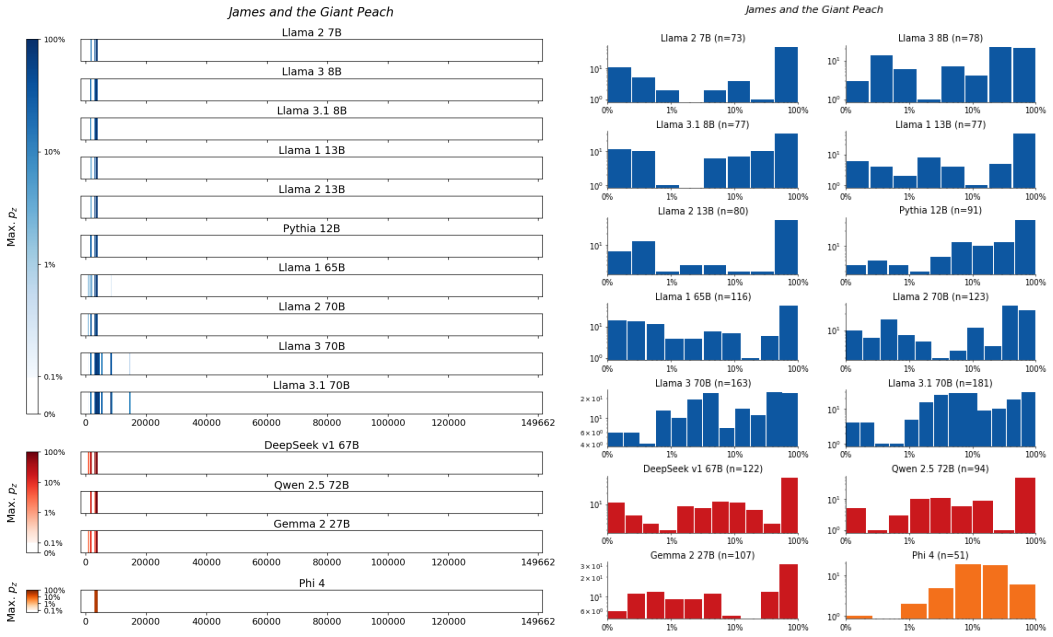


Figure 59: ***James and the Giant Peach,* Dahl.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.49 *Automating the News,* Diakopoulos

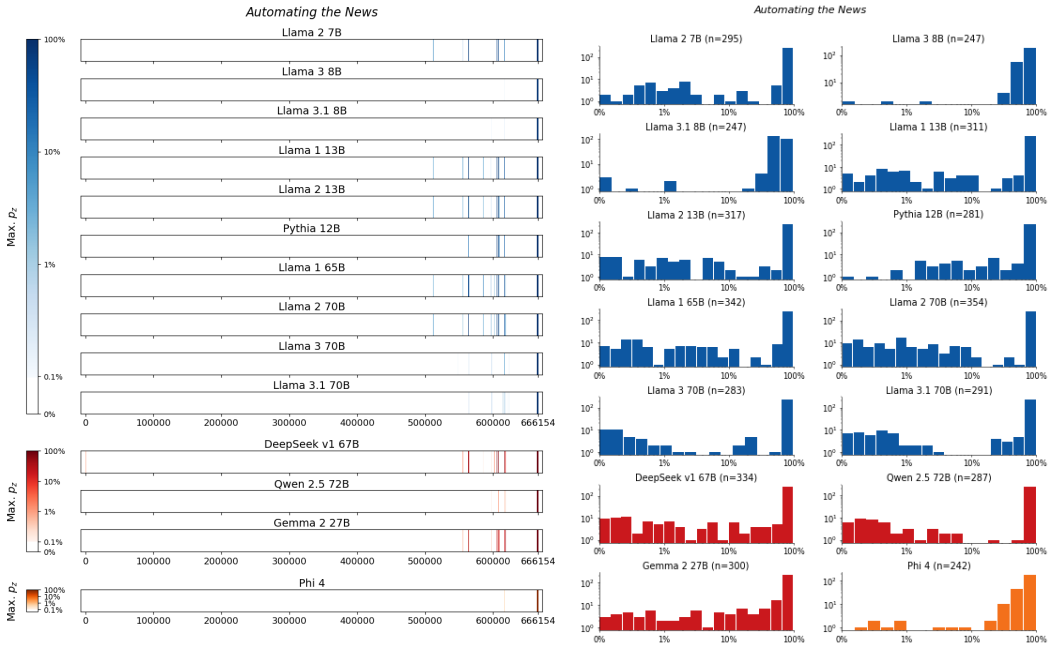


Figure 60: ***Automating the News,* Diakopoulos.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\min} = 0.1\%$).

E.3.50 *Drown,* Díaz

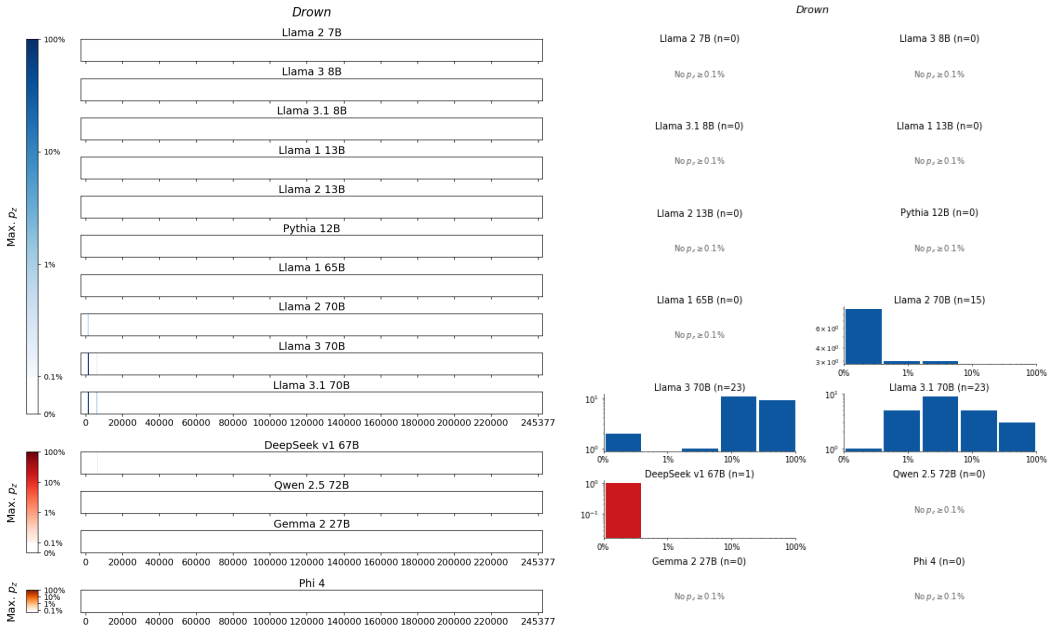


Figure 61: ***Drown,* Díaz.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\min} = 0.1\%$).

E.3.51 *The Brief Wondrous Life of Oscar Wao, Díaz*

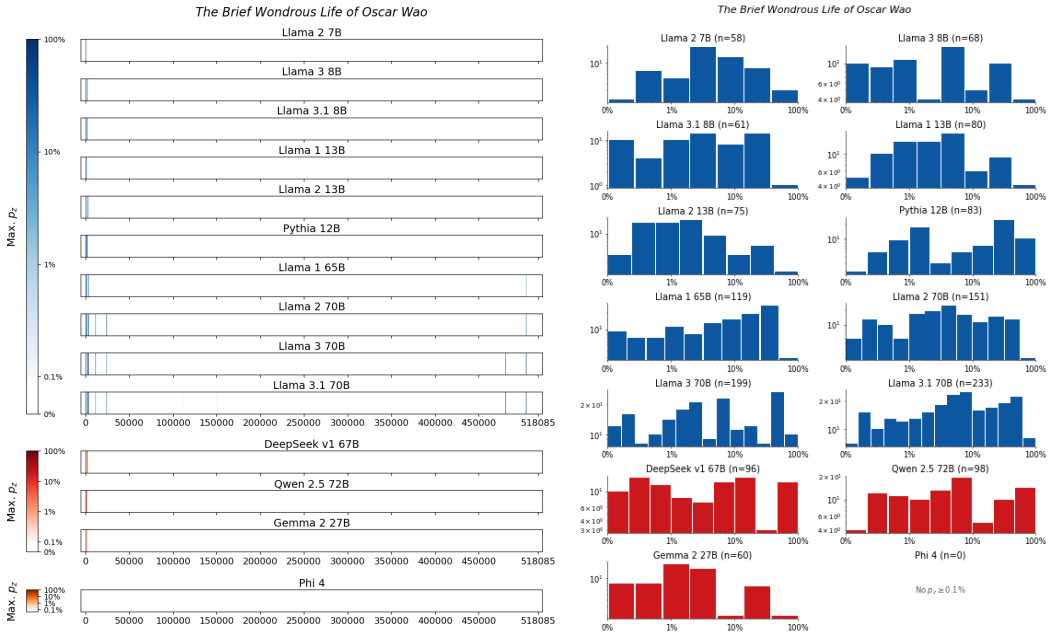


Figure 62: ***The Brief Wondrous Life of Oscar Wao, Díaz.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.52 *This Is How You Lose Her, Díaz*

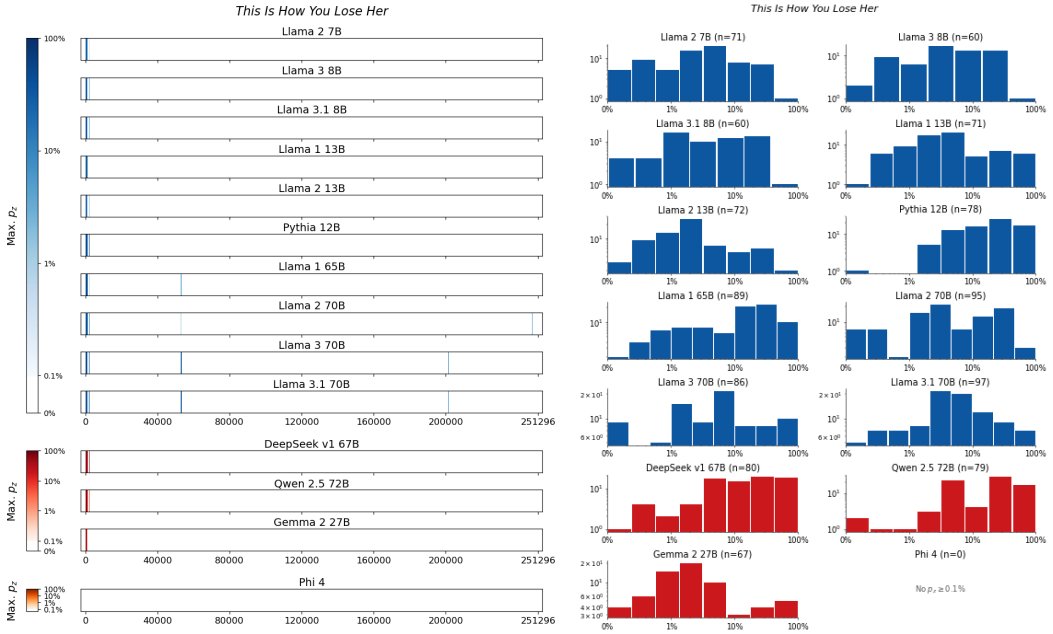


Figure 63: ***This Is How You Lose Her, Díaz.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.53 *The White Album,* Didion

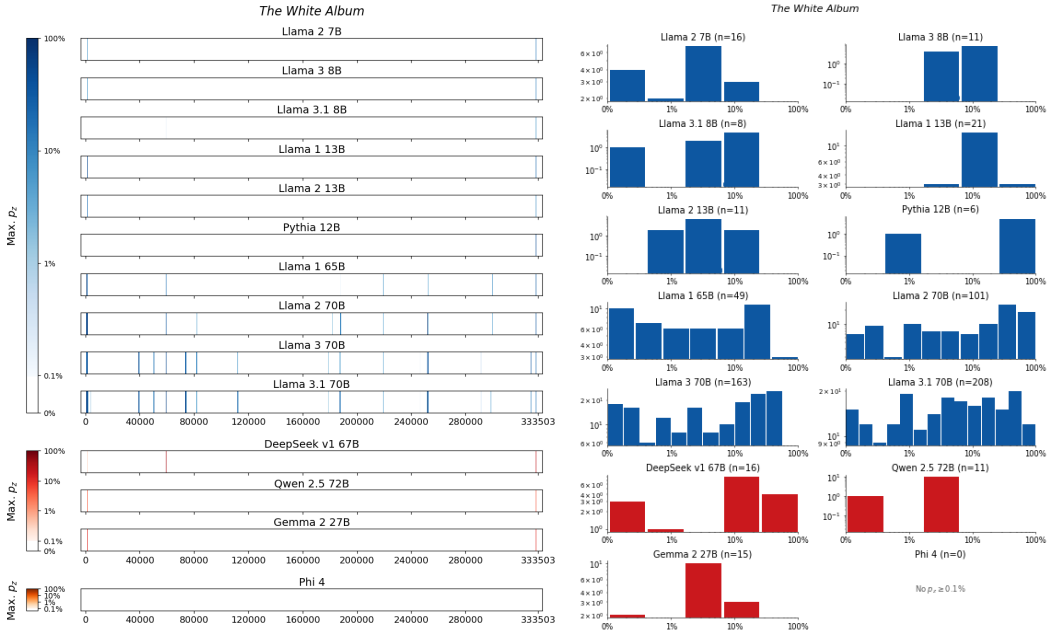


Figure 64: ***The White Album,* Didion.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.54 *Down and Out in the Magic Kingdom,* Doctorow

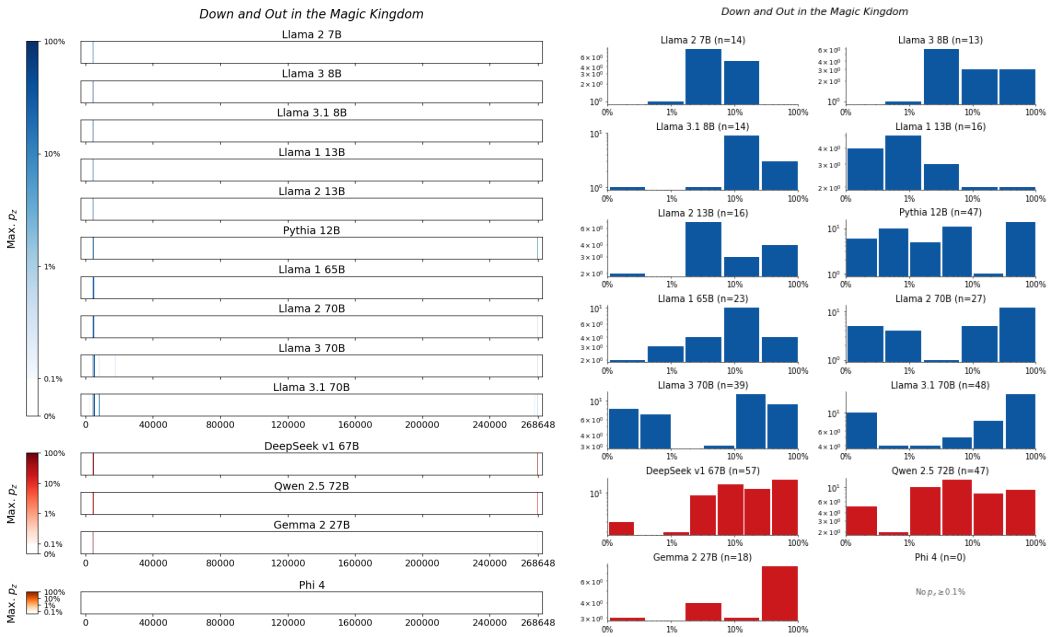


Figure 65: ***Down and Out in the Magic Kingdom,* Doctorow.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.55 *The World's Wife*, *Duffy*

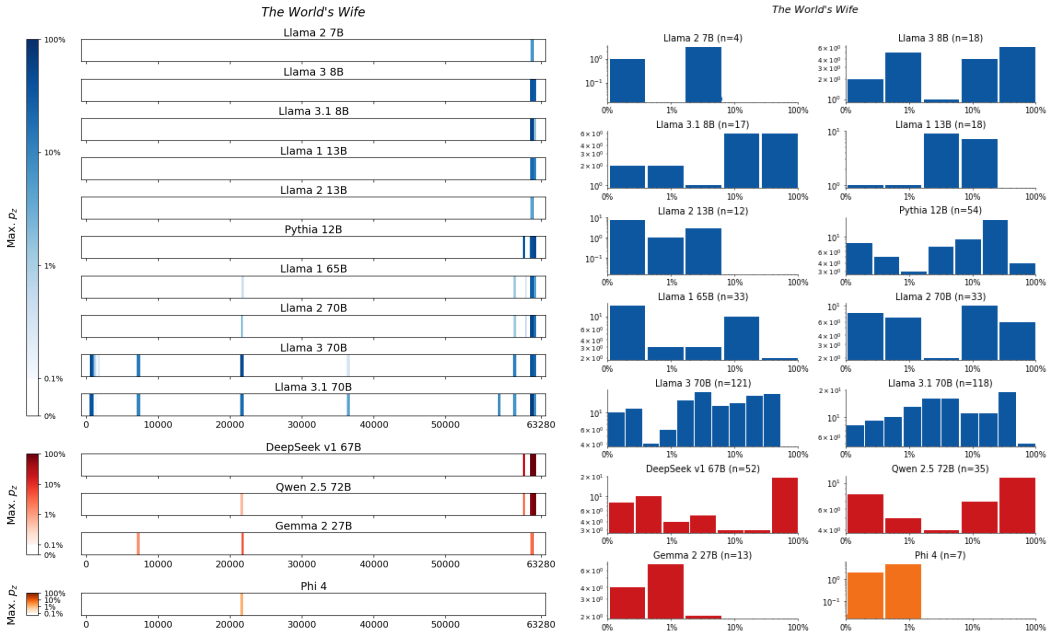


Figure 66: ***The World's Wife*, Duffy.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.56 *A Visit from the Goon Squad*, *Egan*

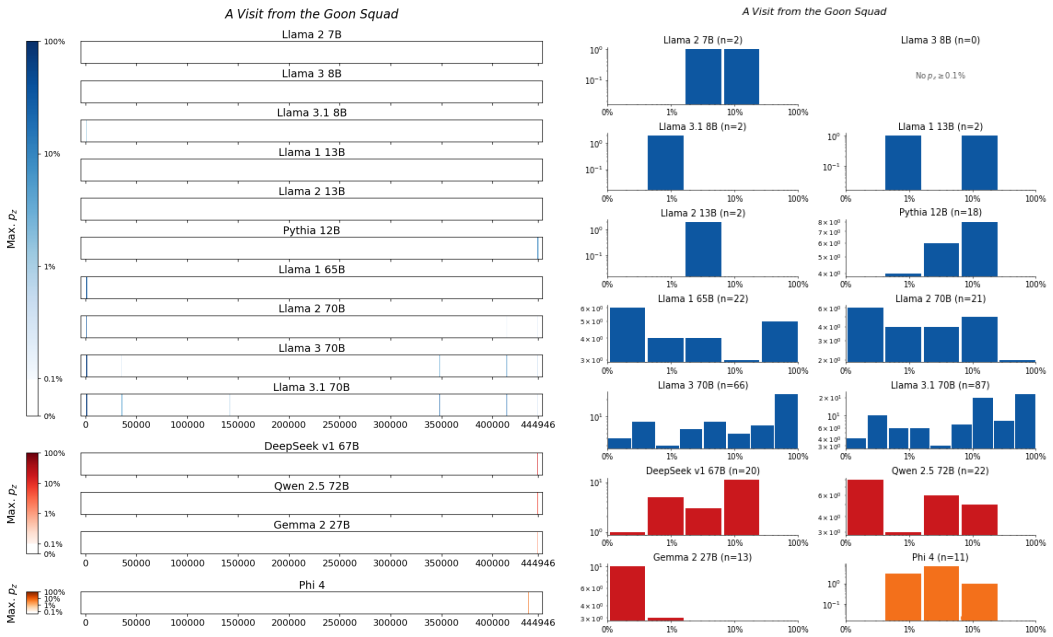


Figure 67: ***A Visit from the Goon Squad*, Egan.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.57 *Invisible Man,* Ellison

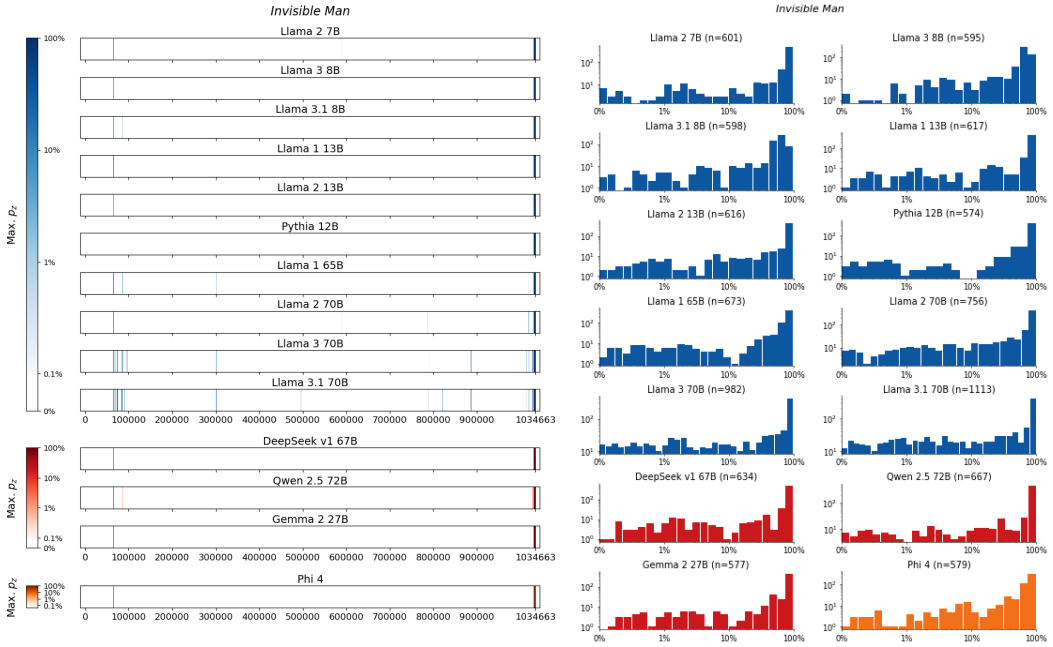


Figure 68: ***Invisible Man,* Ellison.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.58 *We Should All Be Mirandas,* Fairless & Garroni

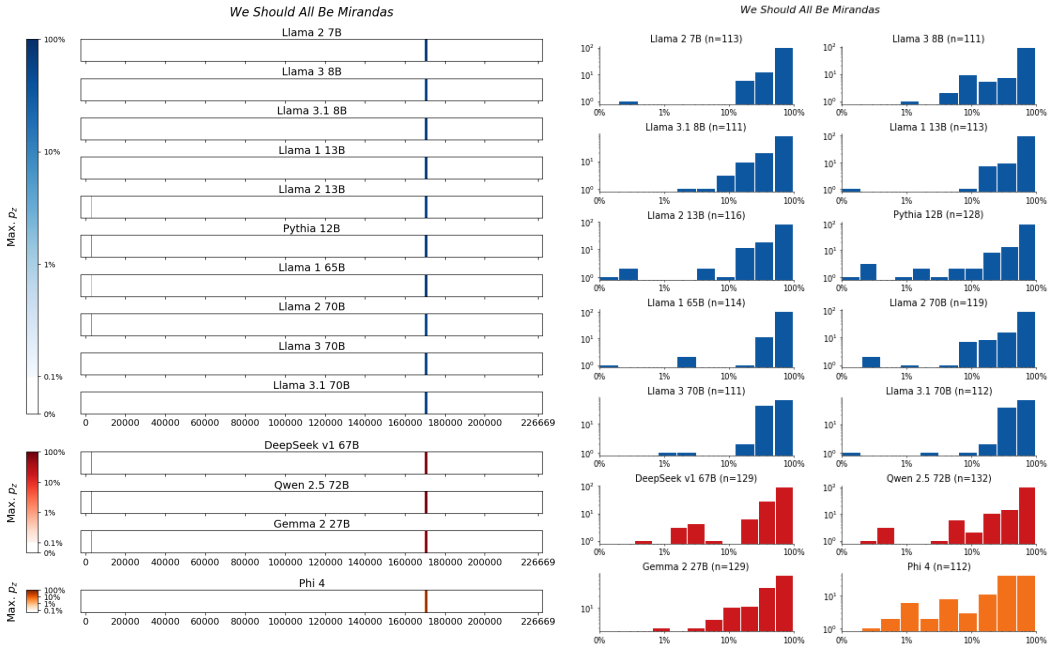


Figure 69: ***We Should All Be Mirandas,* Fairless & Garroni.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.59 *The Dude Abides,* Falsani

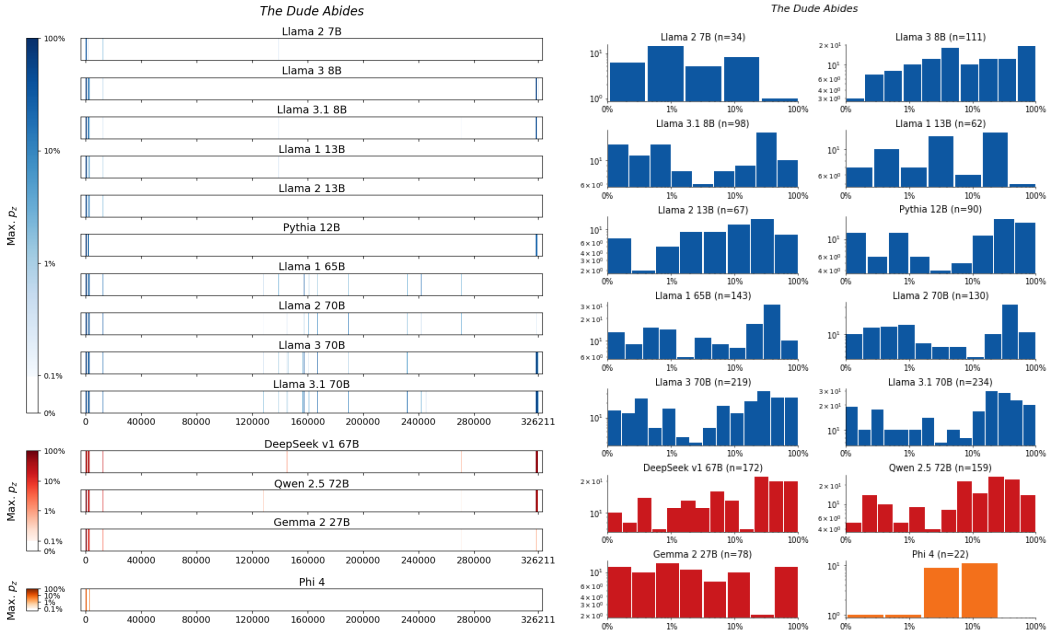


Figure 70: ***The Dude Abides*, Falsani.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.60 *The President's Vampire,* Farnsworth

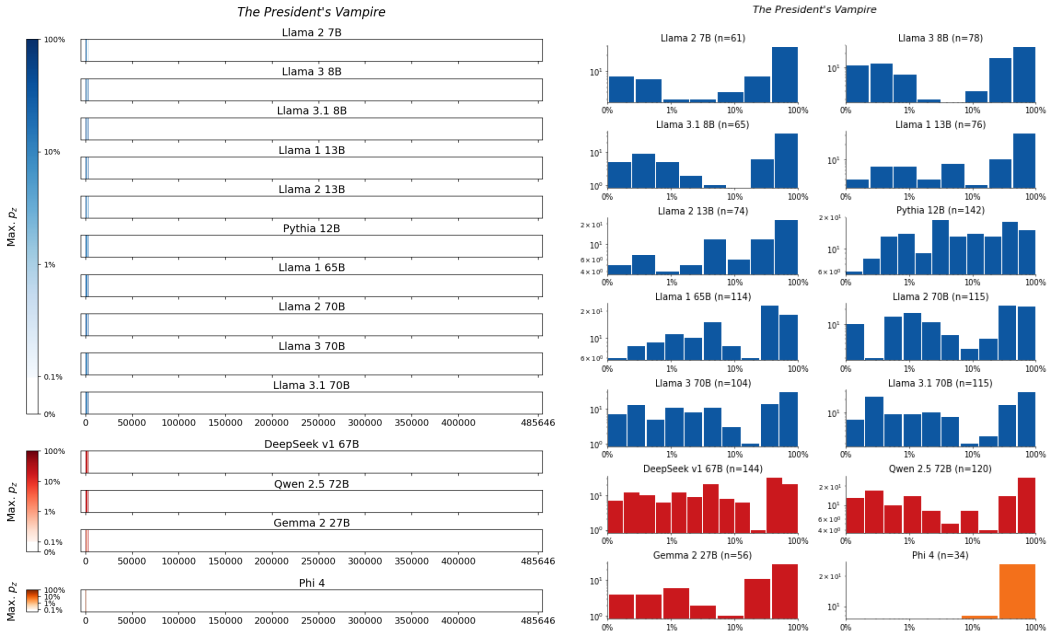


Figure 71: ***The President's Vampire*, Farnsworth.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.61 *The Great Gatsby*, *Fitzgerald*

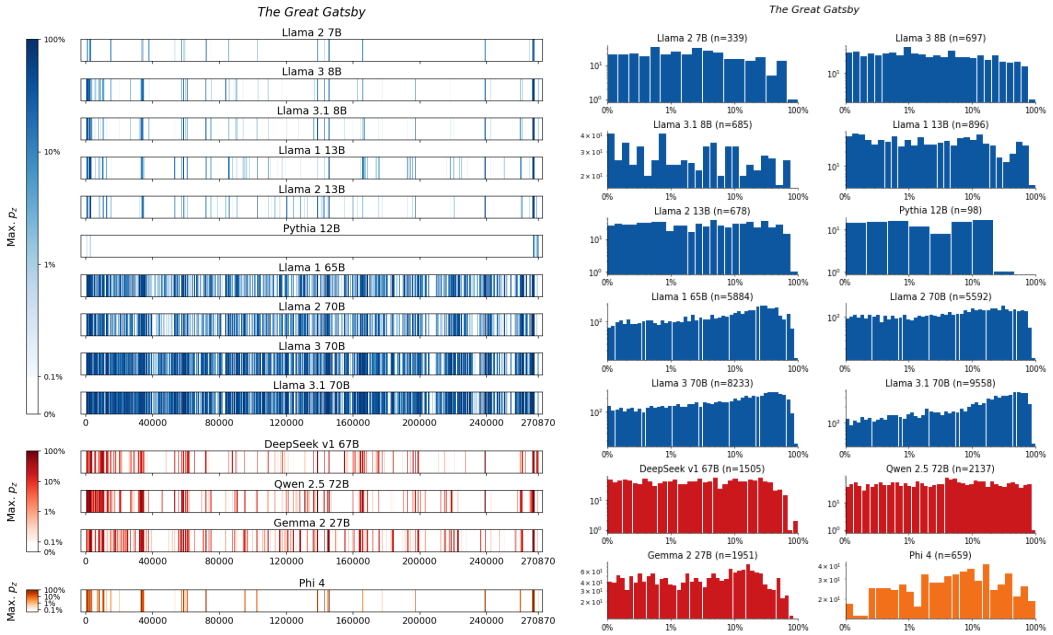


Figure 72: ***The Great Gatsby*, Fitzgerald.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$). This is a public domain book, so PHI 4 is not a negative control.

E.3.62 *Gone Girl*, *Flynn*

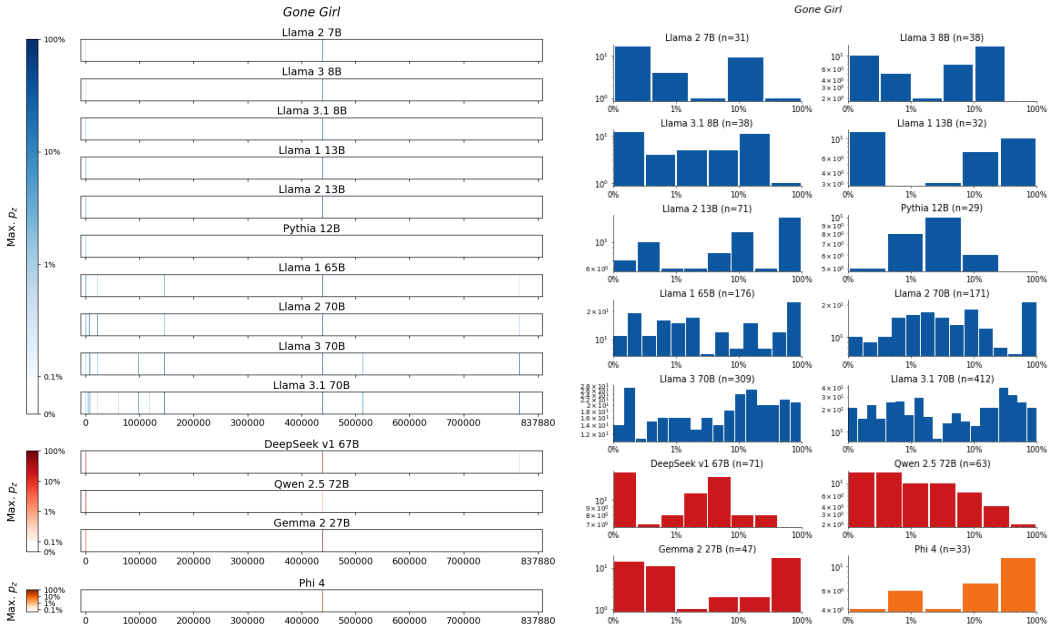


Figure 73: ***Gone Girl*, Flynn.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.63 *British Destroyers,* ***Friedman***

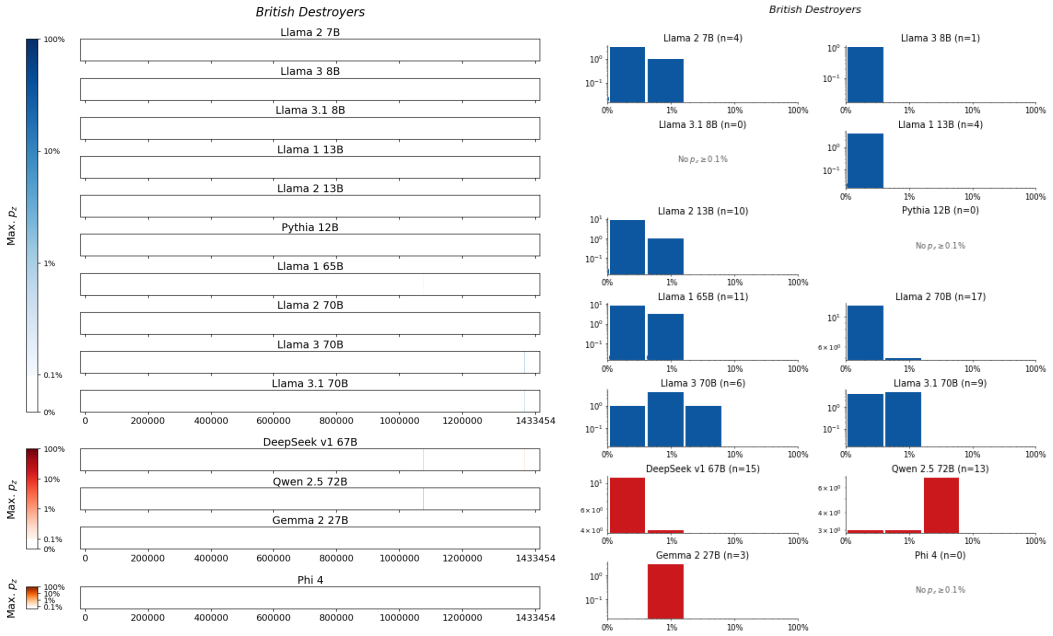


Figure 74: ***British Destroyers*, Friedman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.64 *Florals & Botanicals,* ***Fudurich et al.***

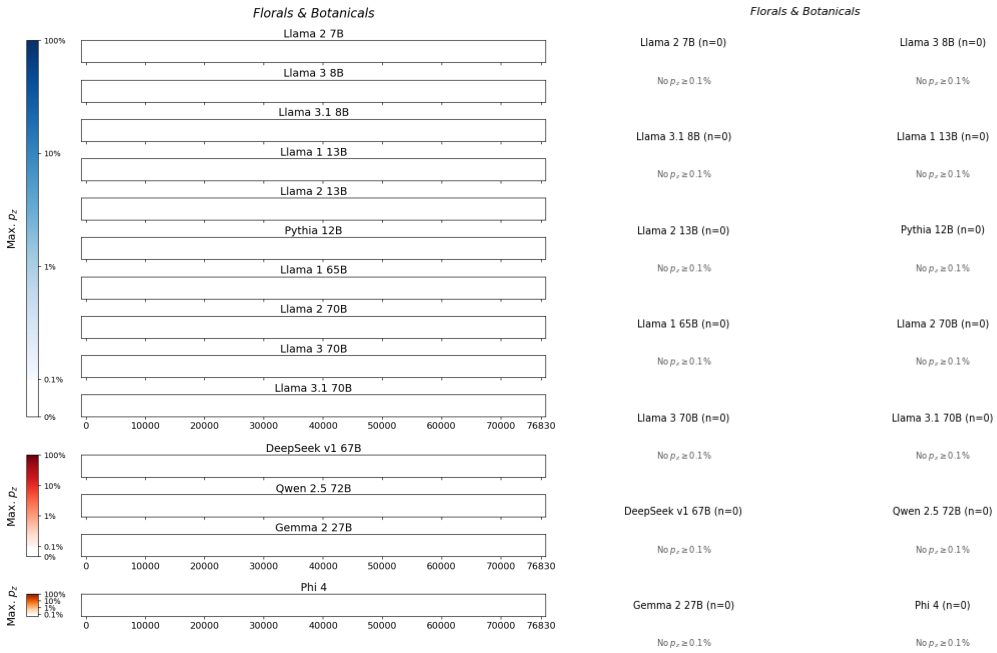


Figure 75: ***Florals & Botanicals*, Fudurich et al..** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.65 *Good Omens, Gaiman & Pratchett*

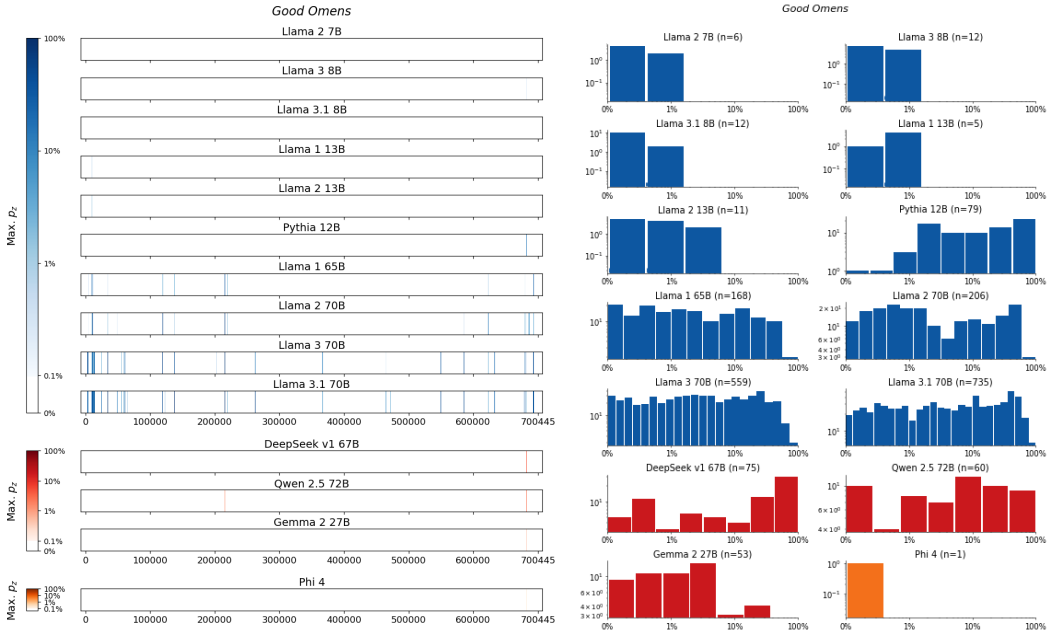


Figure 76: ***Good Omens*, Gaiman & Pratchett.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.66 *The Slippery Year, Gideon*

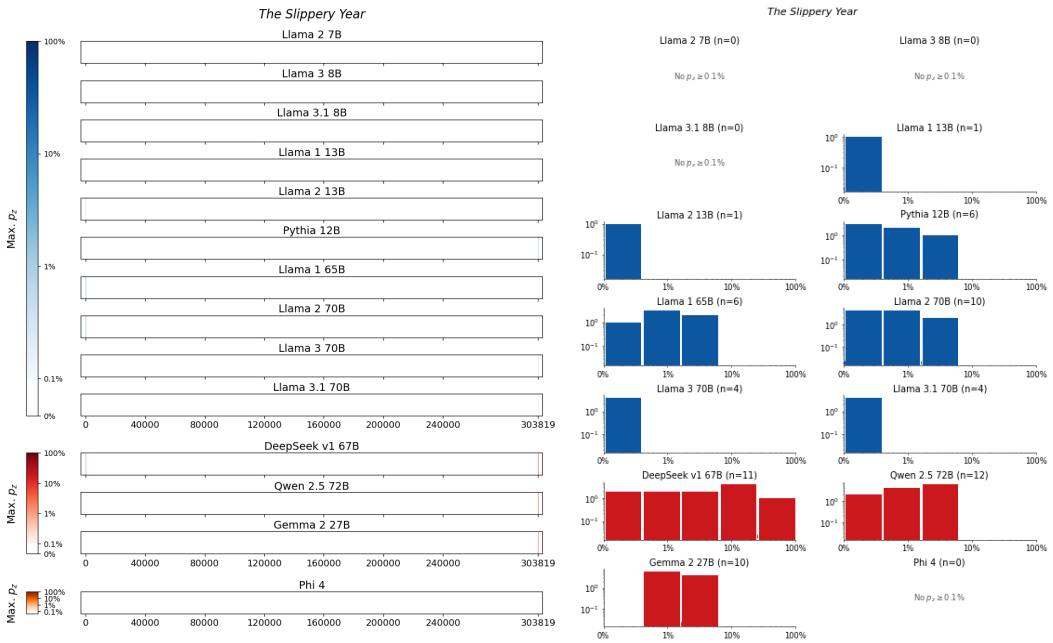


Figure 77: ***The Slippery Year*, Gideon.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.67 *Sweater Surgery, Girard*

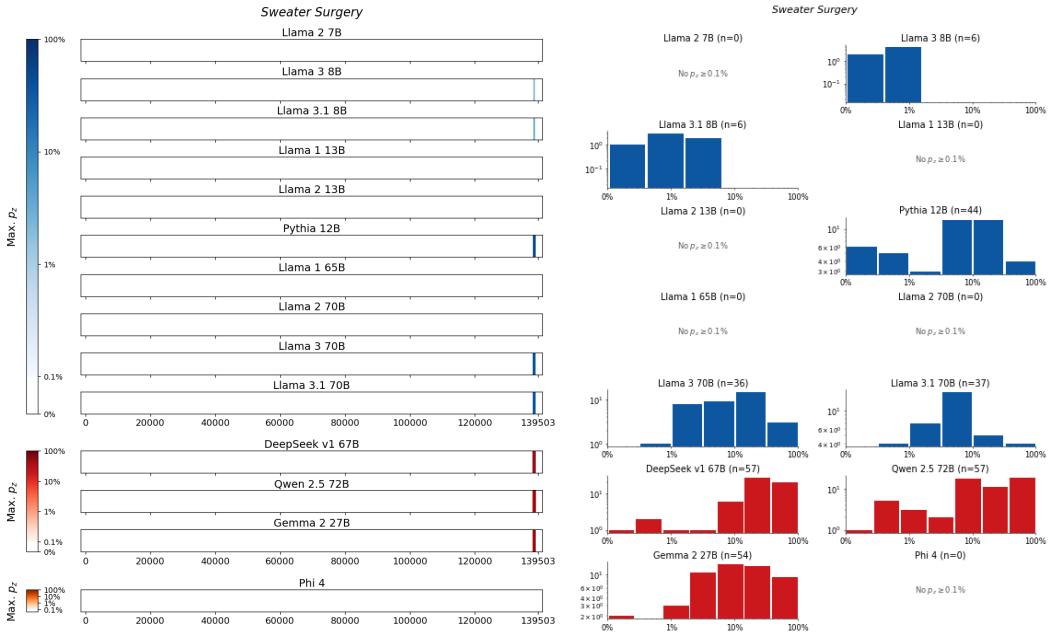


Figure 78: ***Sweater Surgery*, Girard.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.68 *Blink, Gladwell*

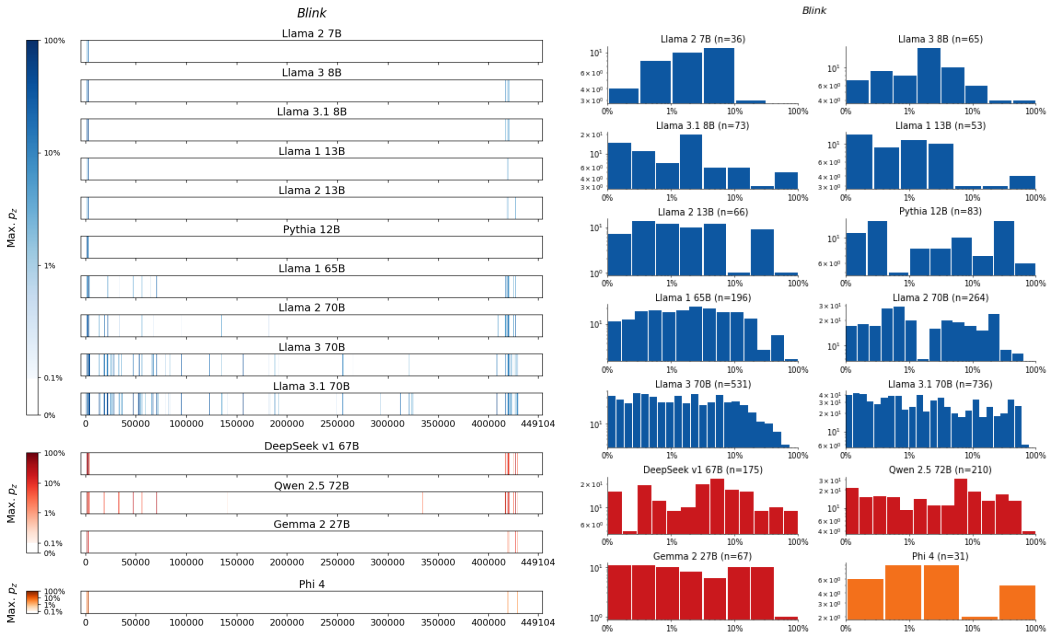


Figure 79: ***Blink*, Gladwell.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.69 *The Land Before Avocado,* Glover

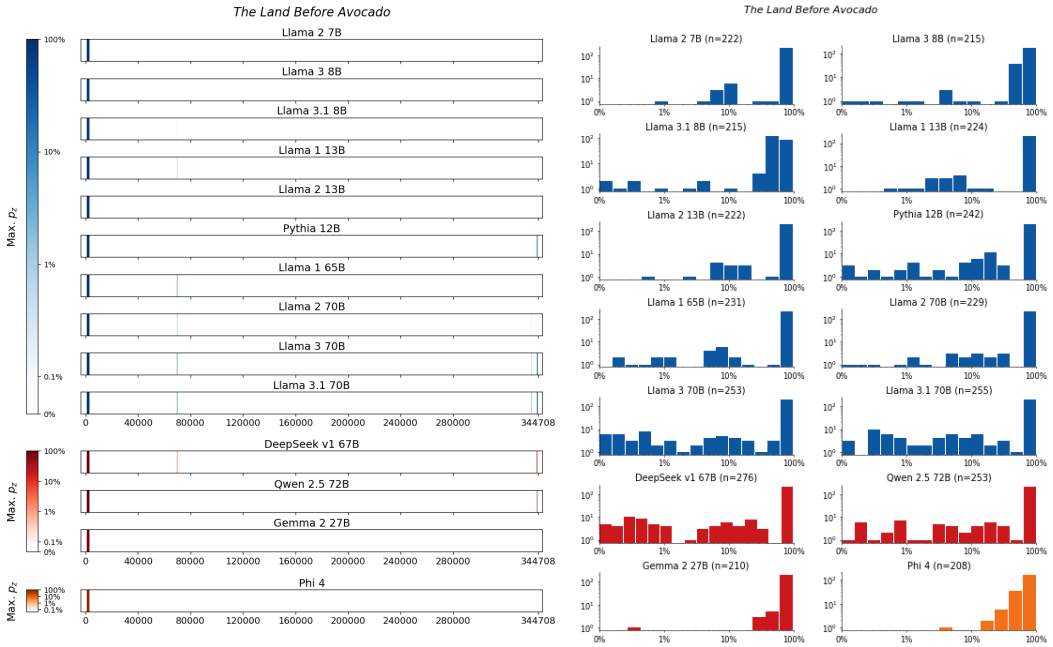


Figure 80: ***The Land Before Avocado,* Glover.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.70 *Dead Ringers,* Golden

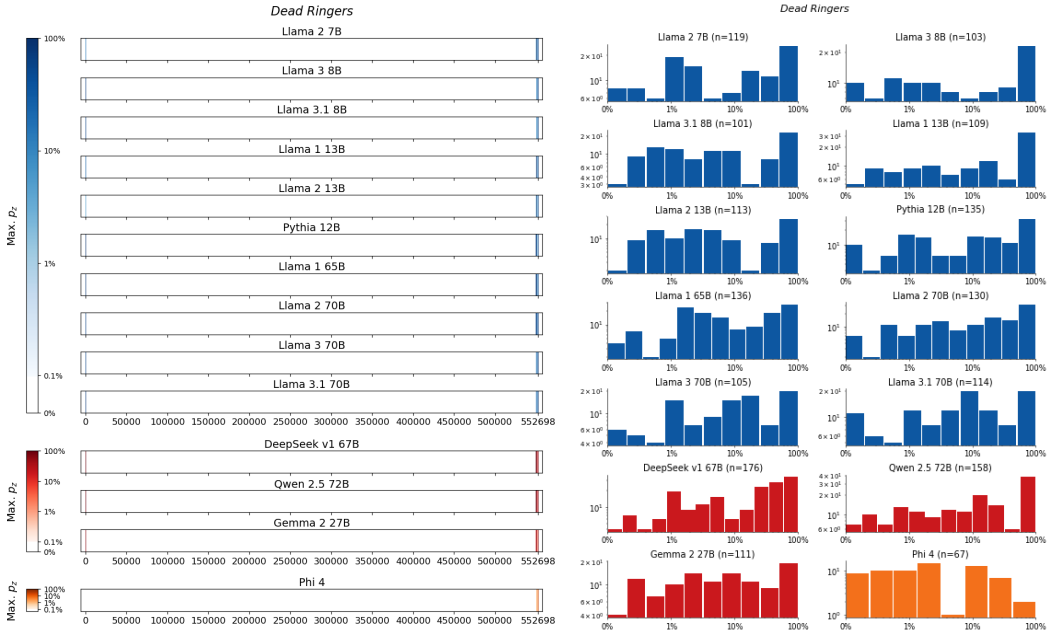


Figure 81: ***Dead Ringers,* Golden.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.71 *Ararat, Golden*

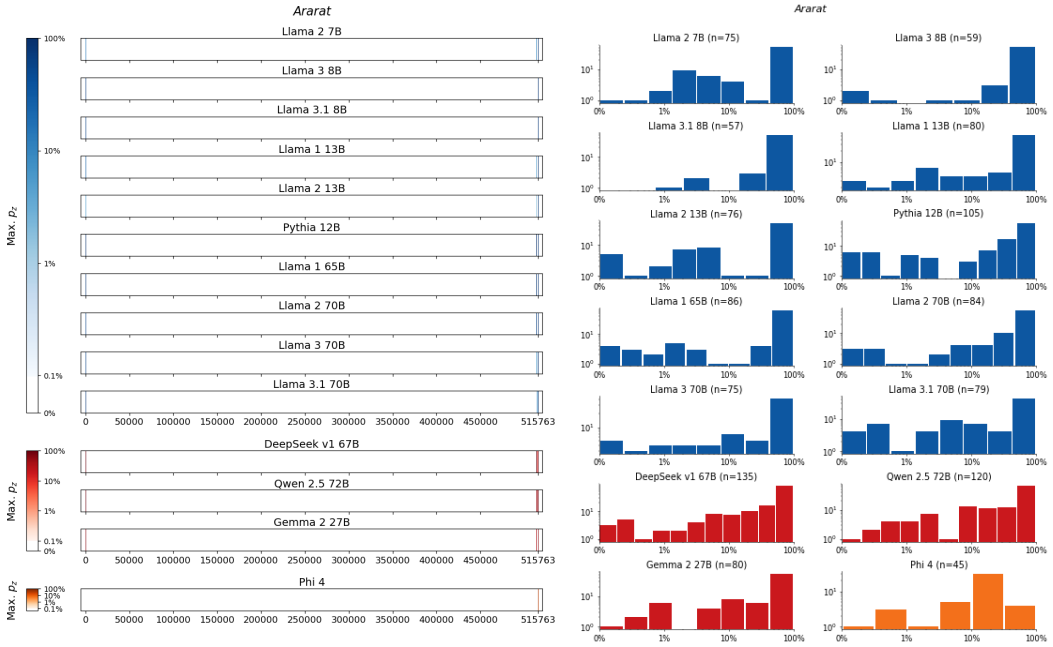


Figure 82: ***Ararat*, Golden.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.72 *Lord of the Flies, Golding*

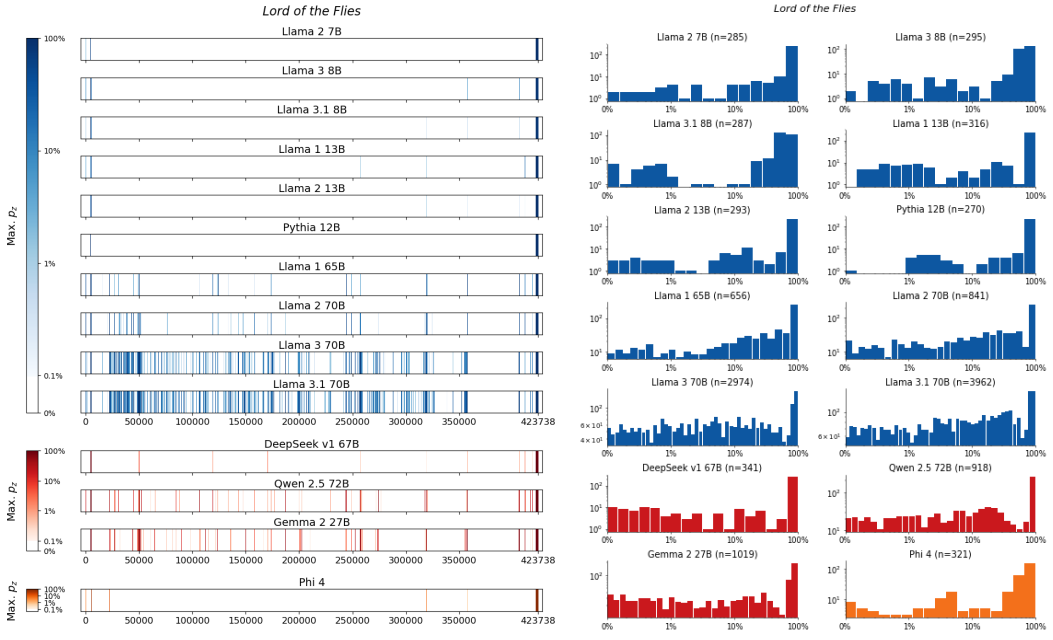


Figure 83: ***Lord of the Flies*, Golding.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.73 *Wizard's First Rule, Goodkind*

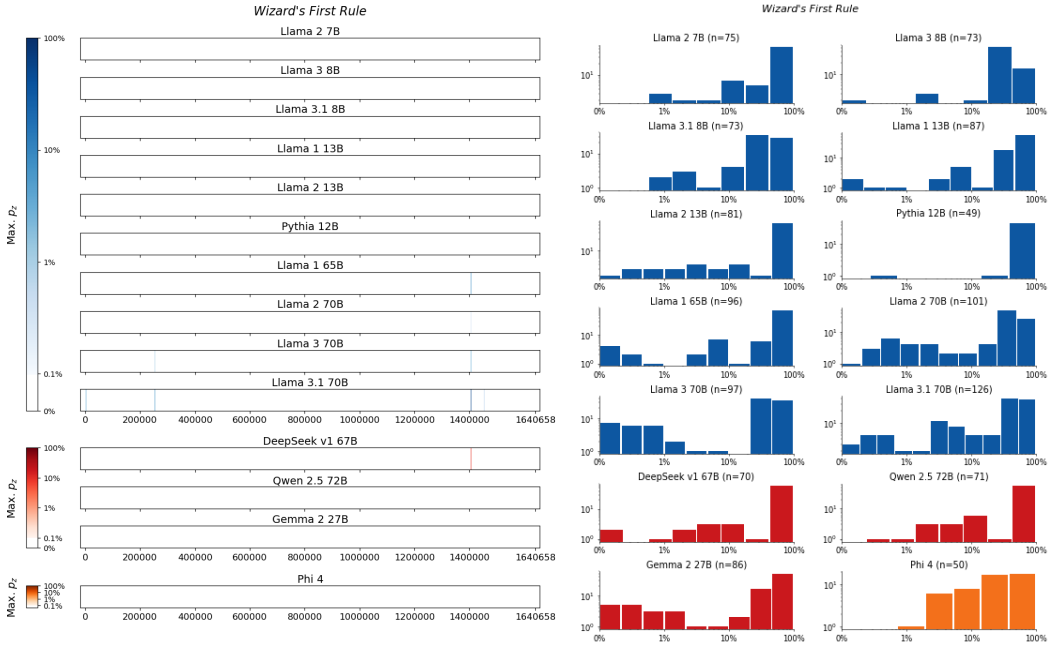


Figure 84: ***Wizard's First Rule*, Goodkind.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.74 *Rome and Jerusalem, Goodman*

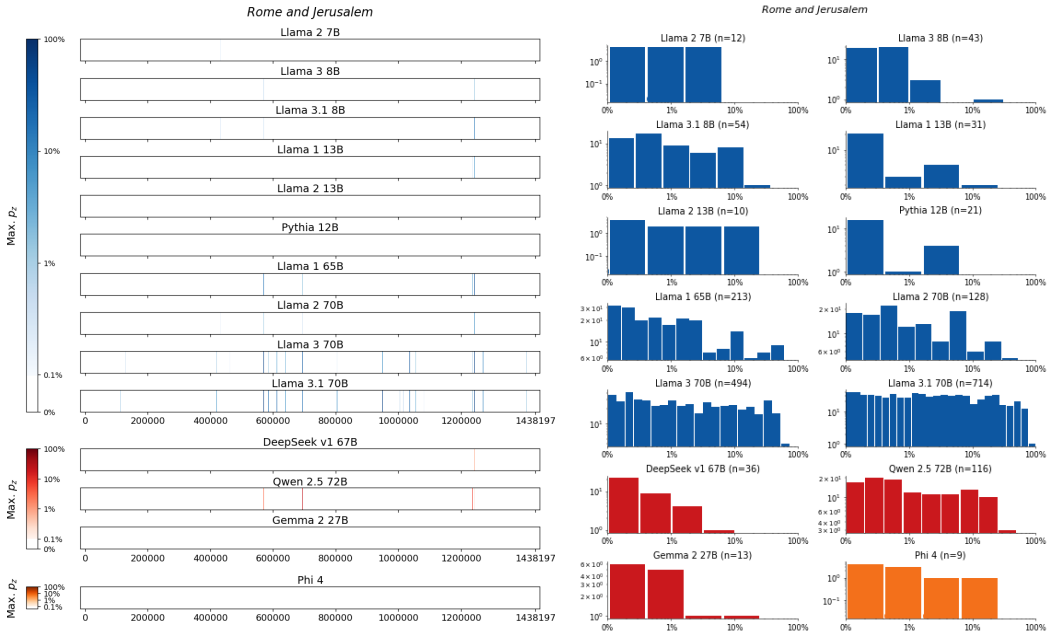


Figure 85: ***Rome and Jerusalem*, Goodman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.75 *The Fault in Our Stars,* Green

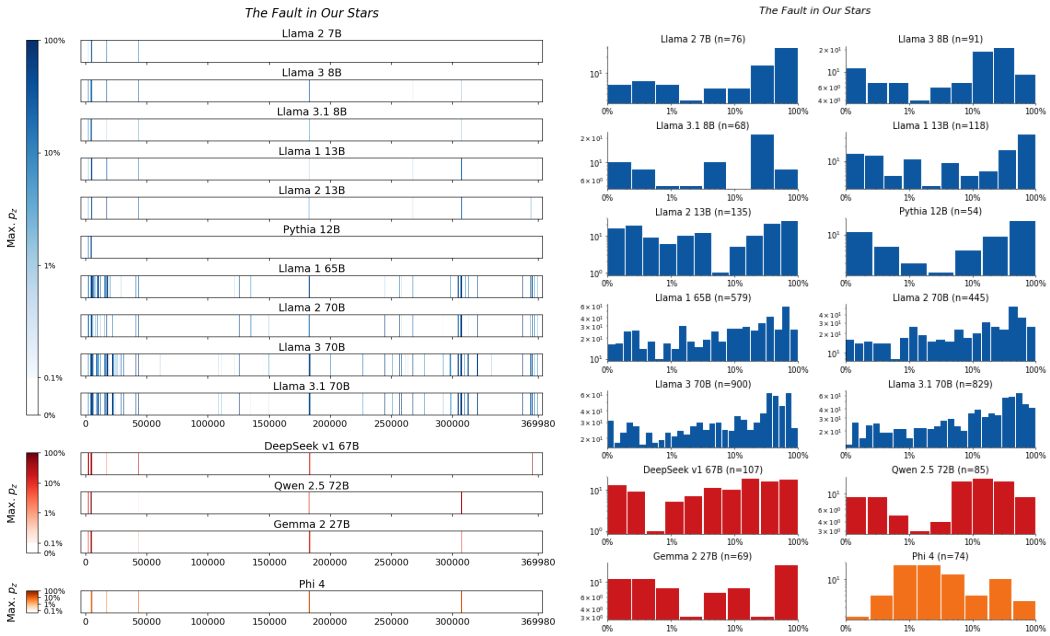


Figure 86: ***The Fault in Our Stars,* Green.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.76 *The Third Man,* Greene

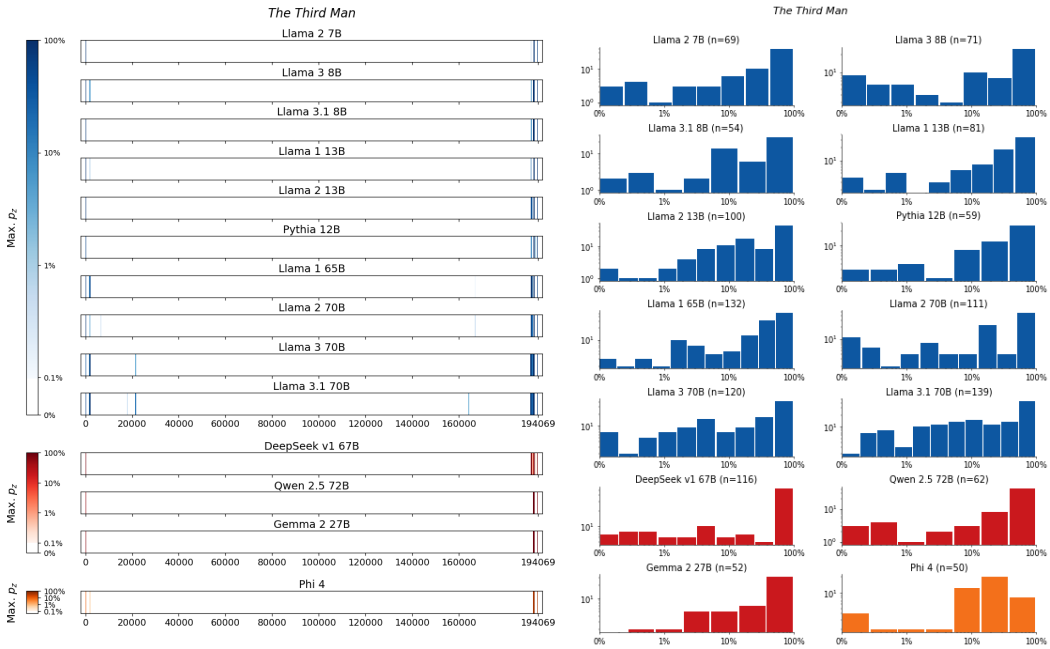


Figure 87: ***The Third Man,* Greene.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.77 *The Confessions of Max Tivoli, Greer*

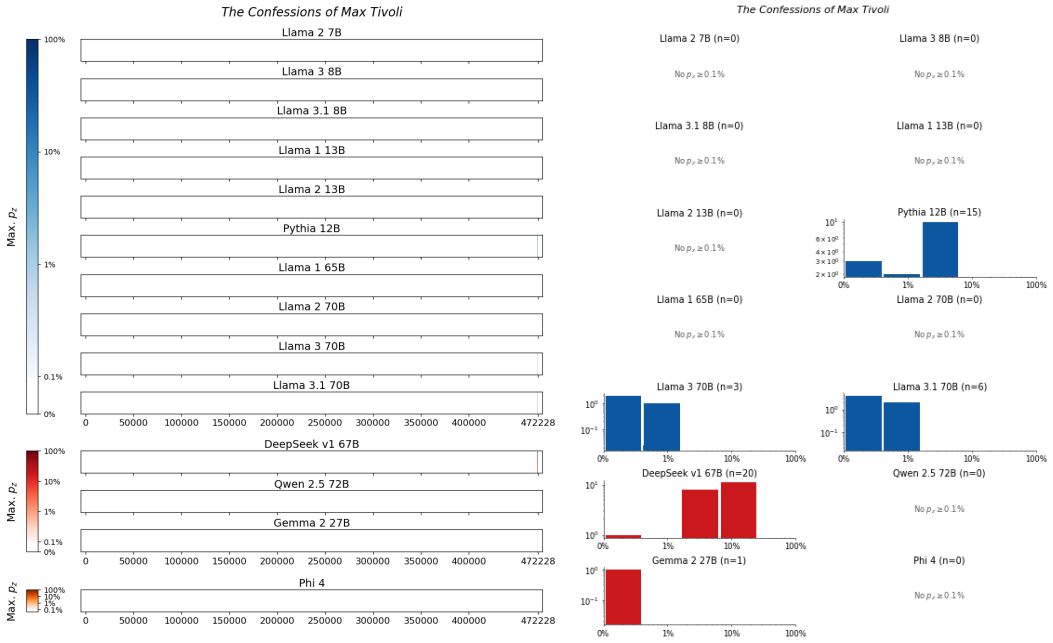


Figure 88: ***The Confessions of Max Tivoli*, Greer.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.78 *The Fugitive, Grisham*

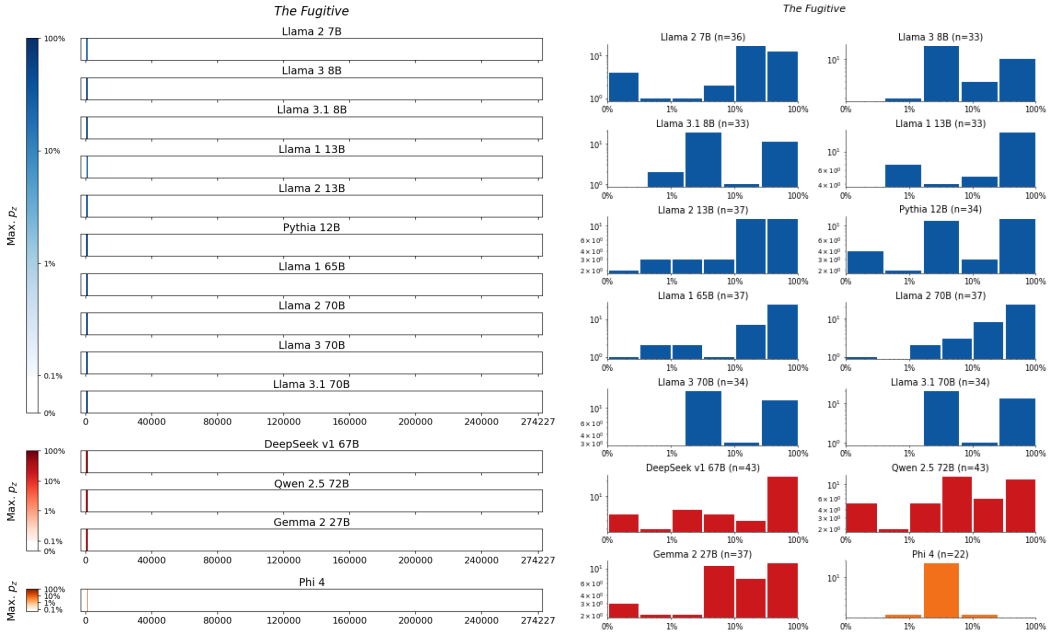


Figure 89: ***The Fugitive*, Grisham.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.79 *The Curious Incident of the Dog in the Night-Time, Haddon*

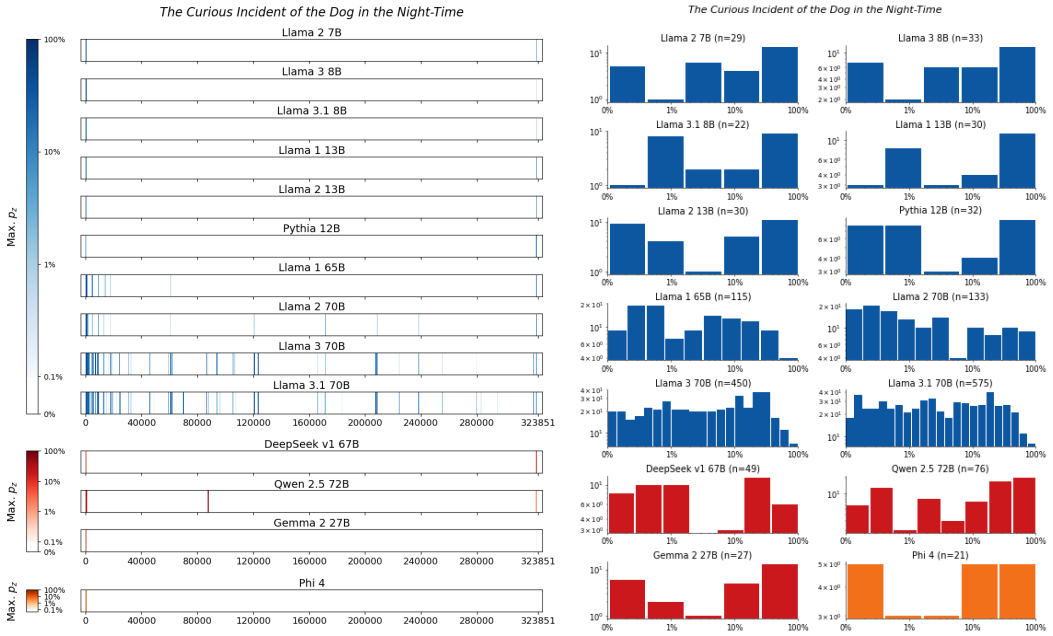


Figure 90: ***The Curious Incident of the Dog in the Night-Time*, Haddon.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.80 *Migrations to Solitude, Halpern*

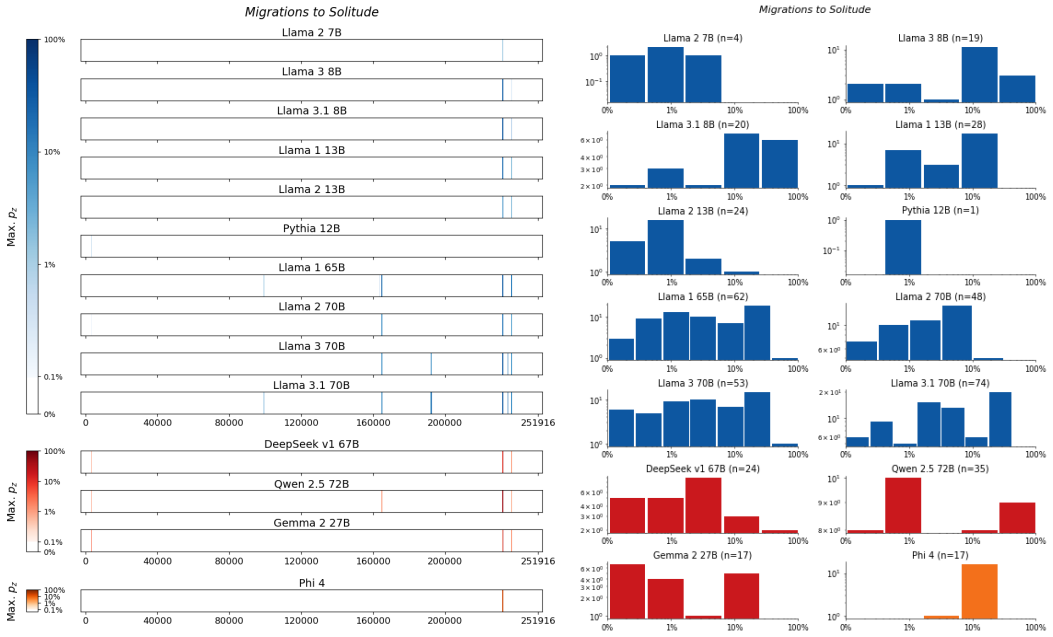


Figure 91: ***Migrations to Solitude*, Halpern.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.81 *Uncommon Type, Hanks*

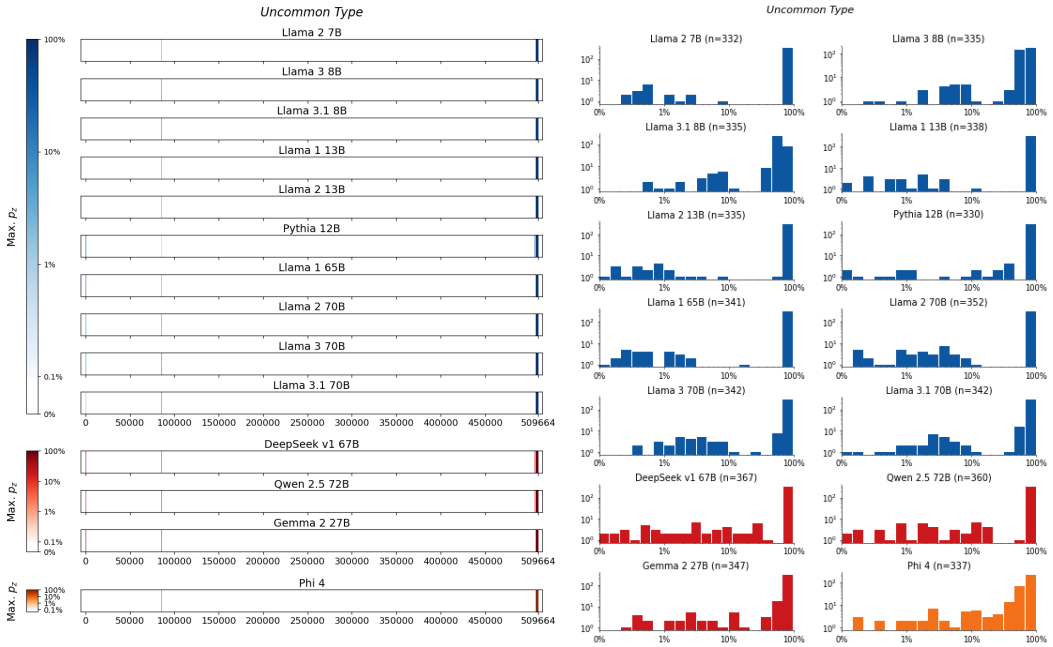


Figure 92: ***Uncommon Type*, Hanks.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.82 *Buzz, Hanson*

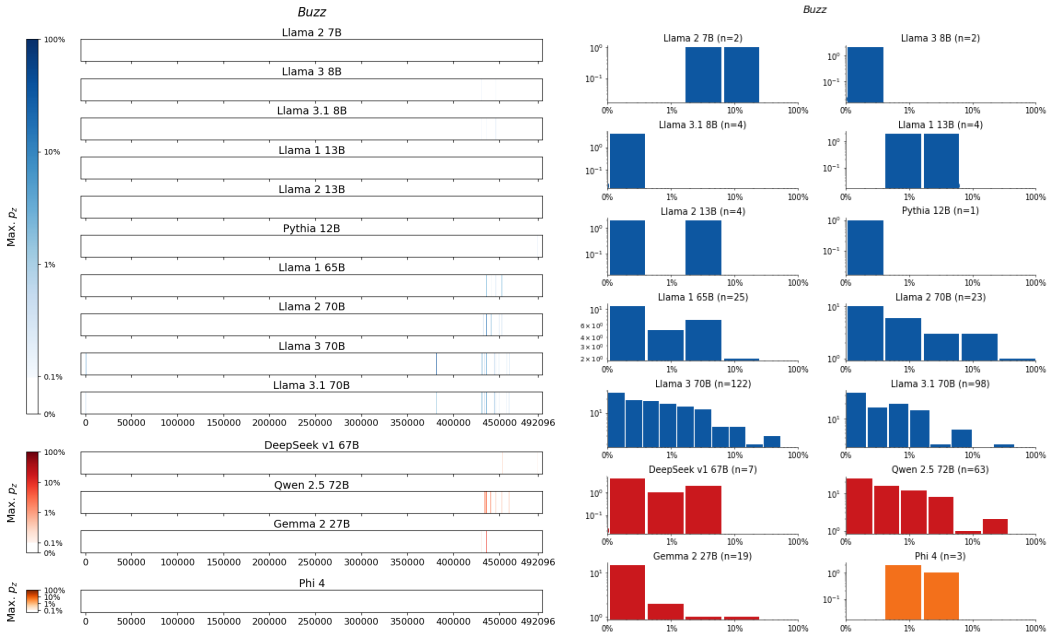


Figure 93: ***Buzz*, Hanson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.83 *Requiem for the Sun,* Haydon

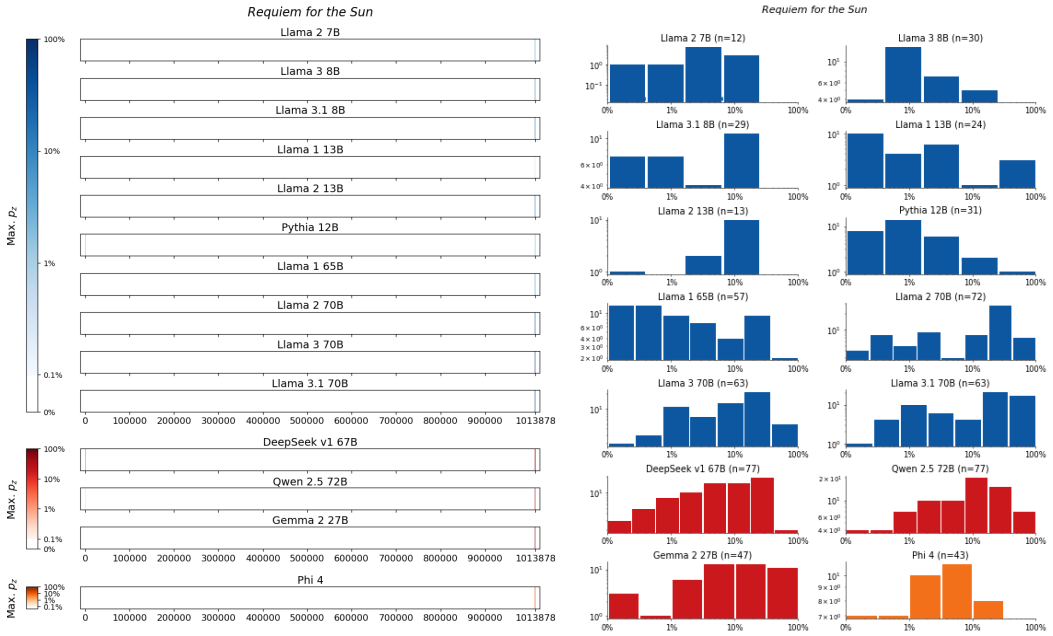


Figure 94: ***Requiem for the Sun,* Haydon.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.84 *Catch-22,* Heller

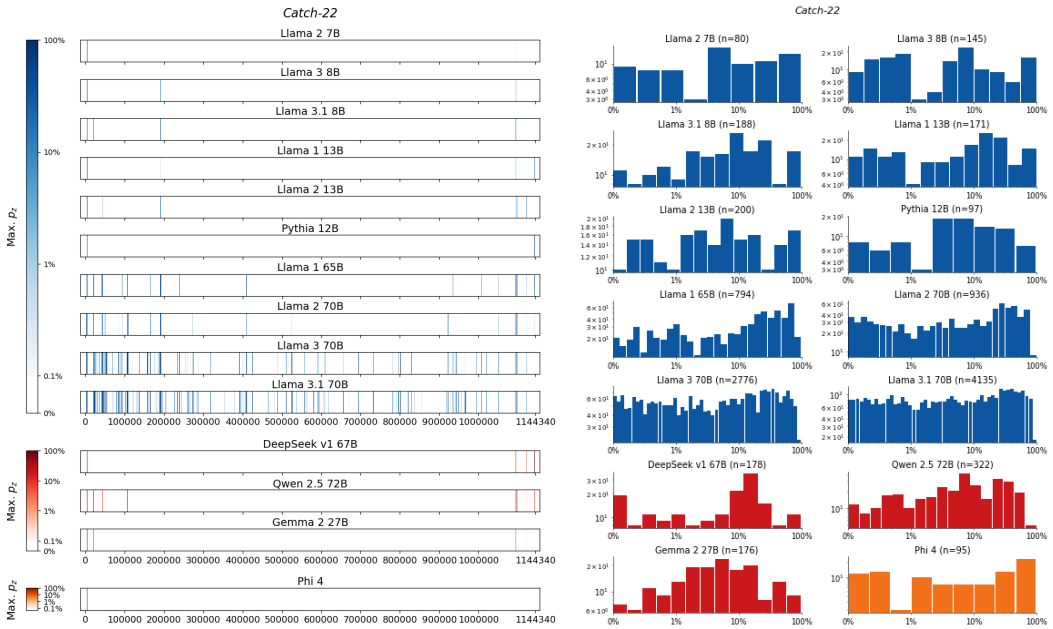


Figure 95: ***Catch-22,* Heller.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.85 *The Old Man and the Sea*, *Hemingway*

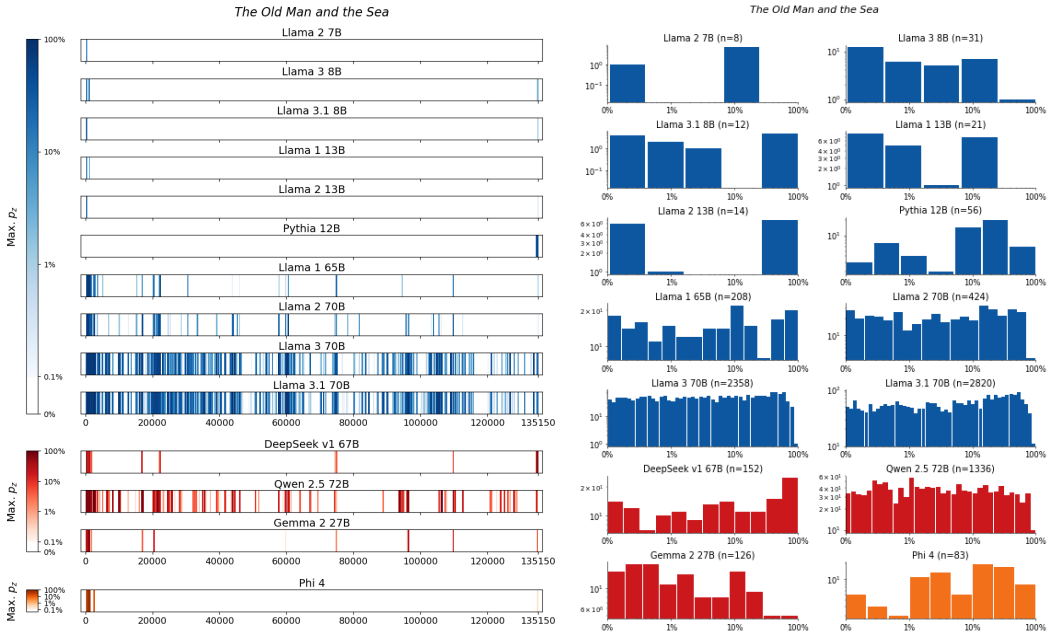


Figure 96: ***The Old Man and the Sea*, Hemingway.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.86 *Life on Air*, *Hendy*

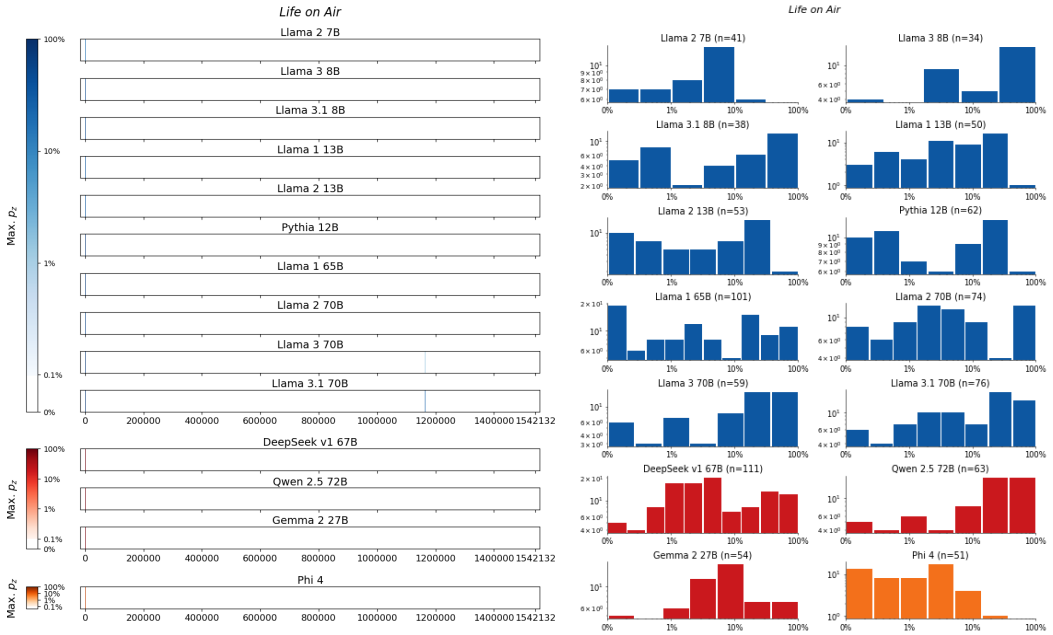


Figure 97: ***Life on Air*, Hendy.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.87 *Great Hair Days,* ***Hersheson***

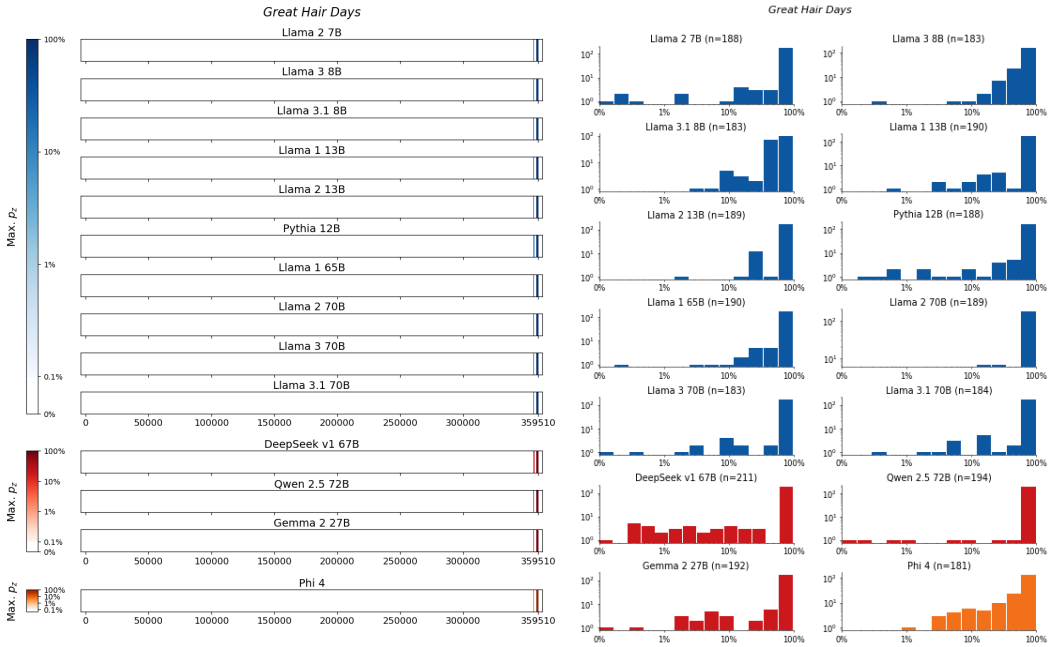


Figure 98: ***Great Hair Days,* Hersheson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.88 *Graft,* ***Hill***

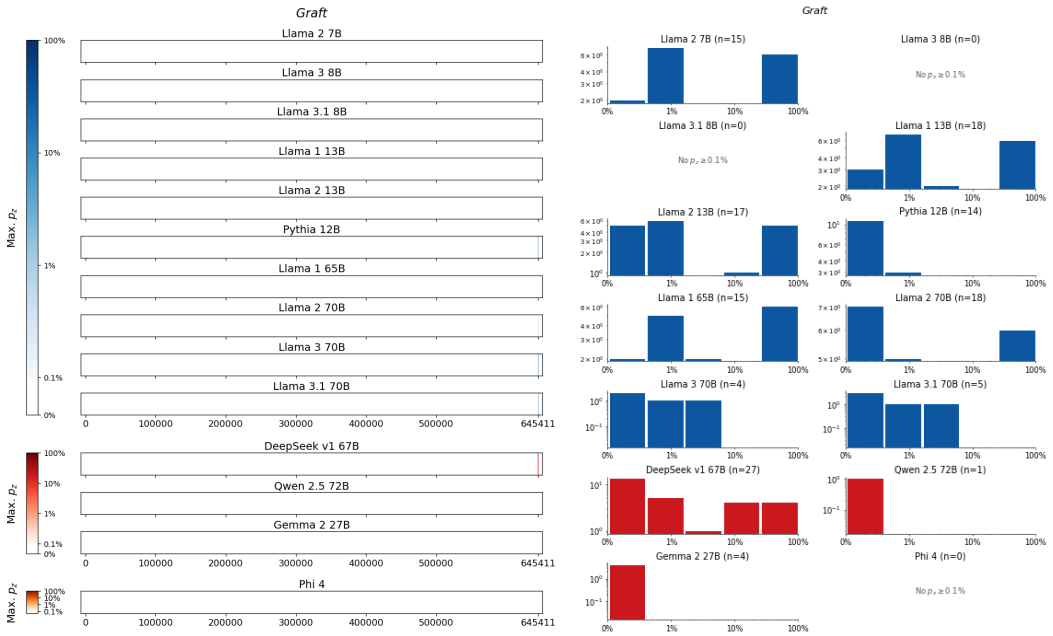


Figure 99: ***Graft,* Hill.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.89 *The Outsiders,* Hinton

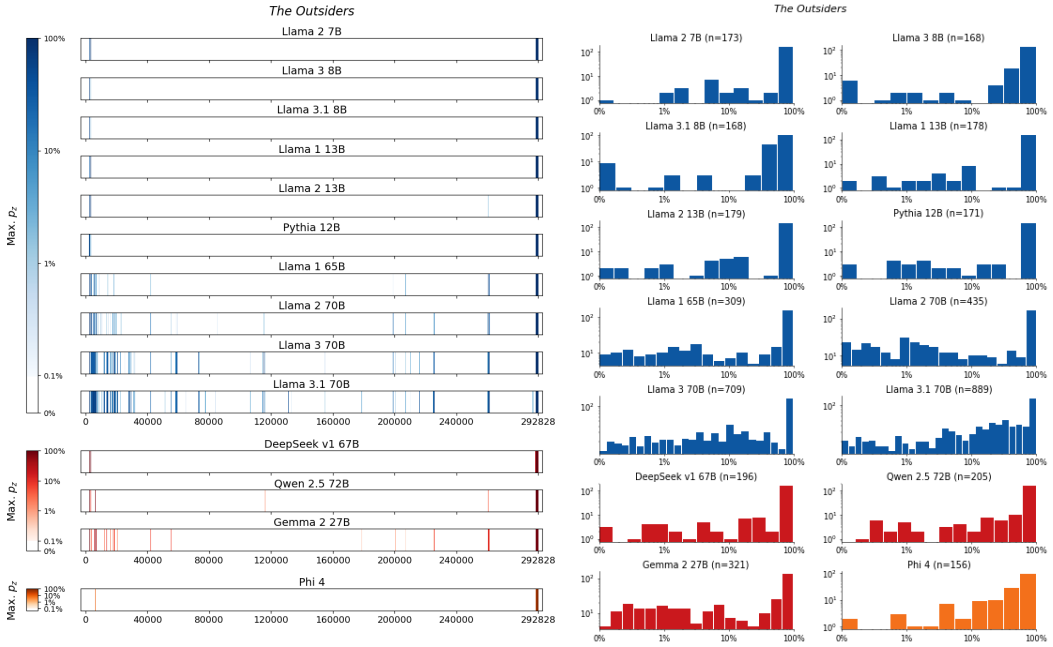


Figure 100: ***The Outsiders,* Hinton.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.90 *The Second Summoning,* Huff

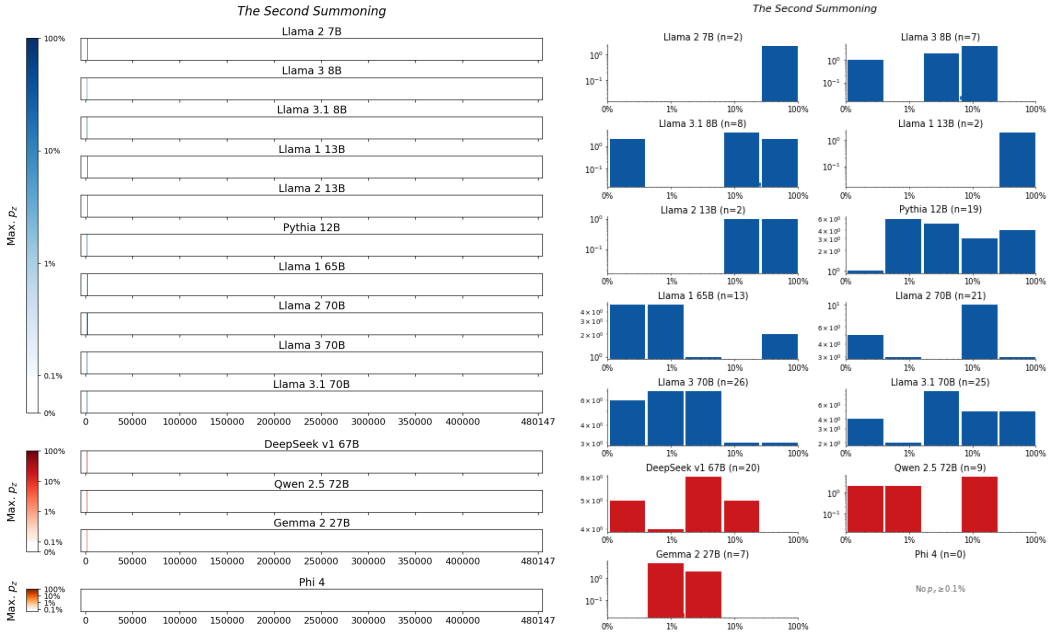


Figure 101: ***The Second Summoning,* Huff.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.91 *Selected Poems of Langston Hughes*, Hughes

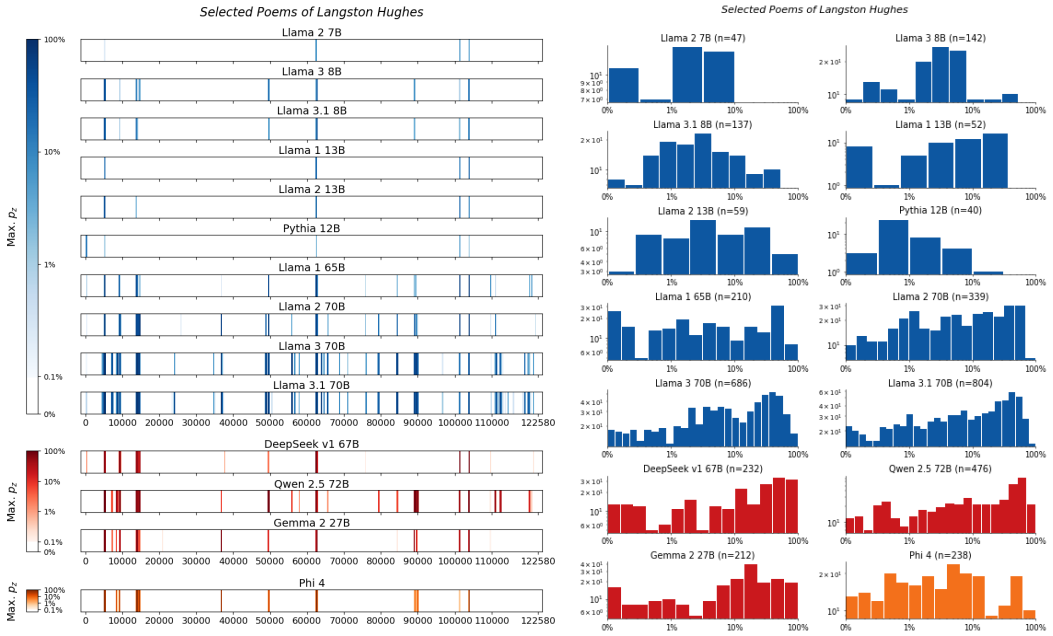


Figure 102: ***Selected Poems of Langston Hughes*, Hughes.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.92 *M. Butterfly*, Hwang

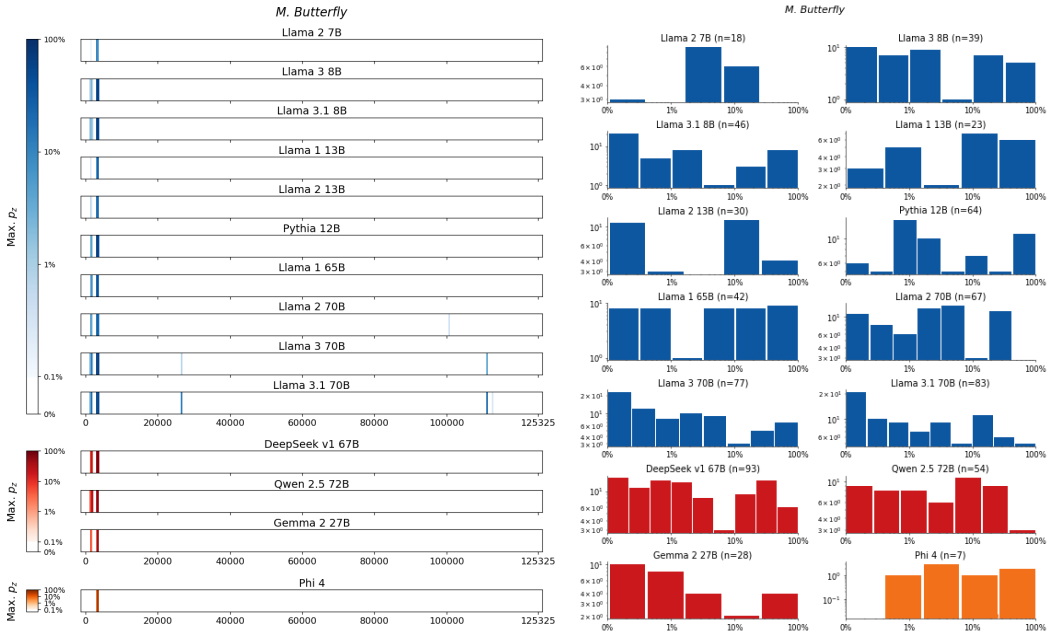


Figure 103: ***M. Butterfly*, Hwang.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.93 *Building and Operating a Realistic Model Railway,* ***Jackson***

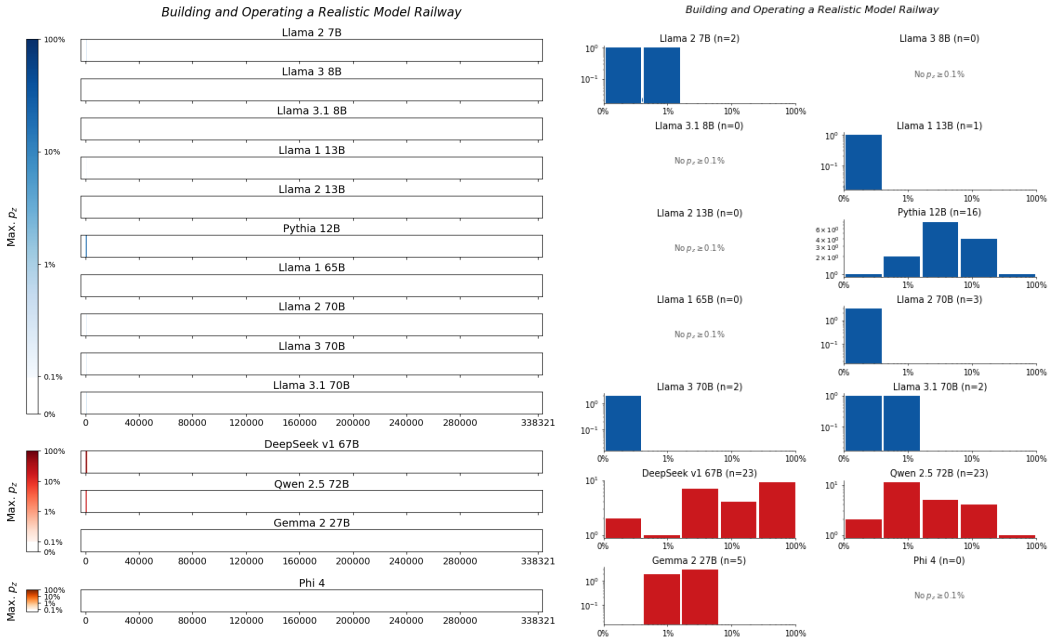


Figure 104: ***Building and Operating a Realistic Model Railway,* Jackson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.94 *All the Onions,* ***Jacobs***

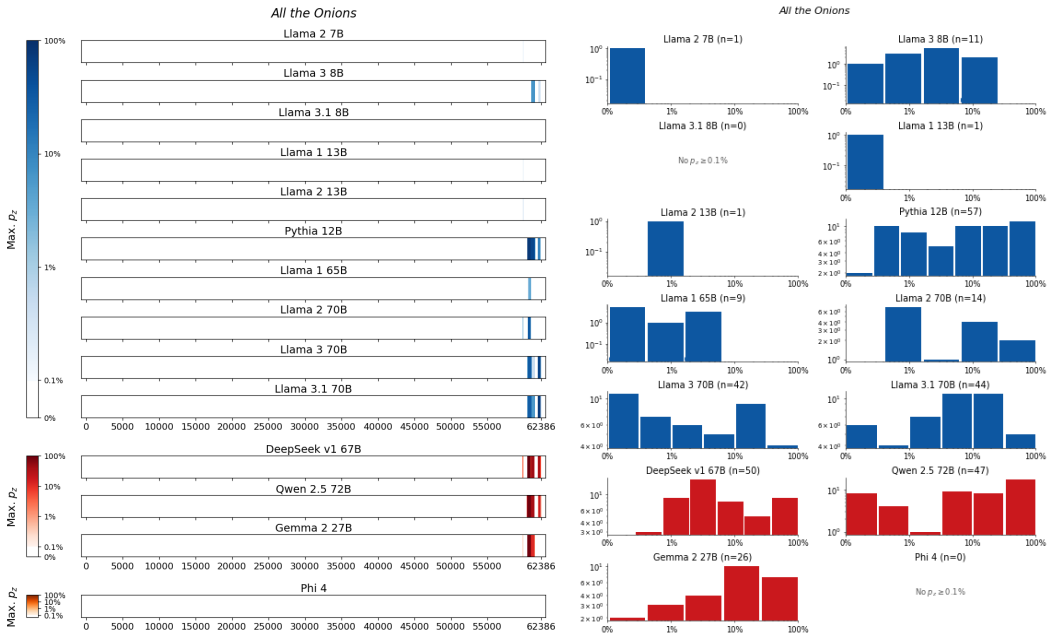


Figure 105: ***All the Onions,* Jacobs.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.95 *Fifty Shades of Grey, James*

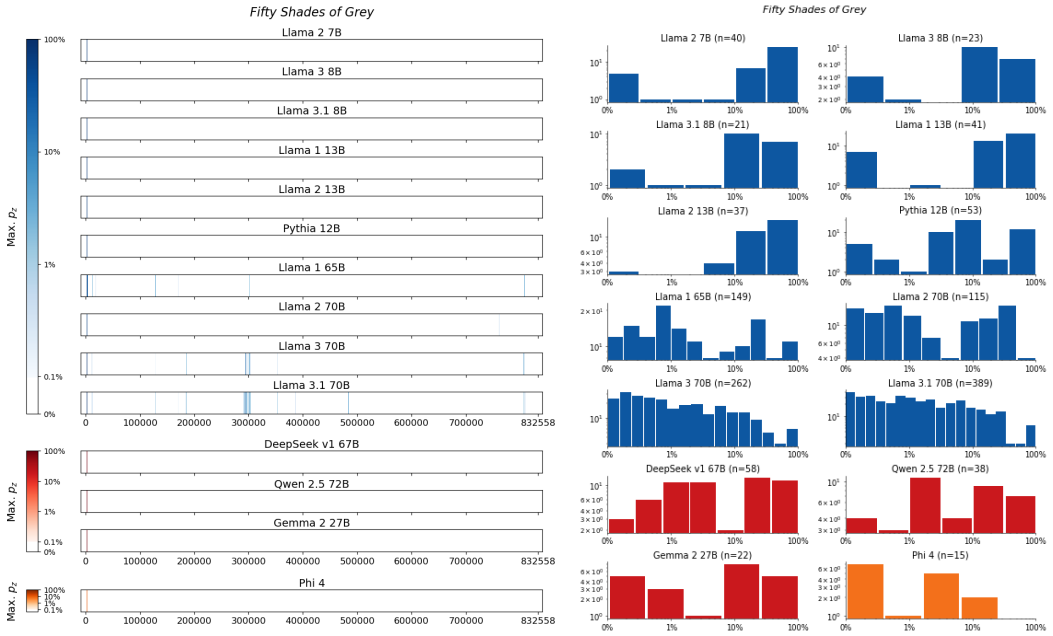


Figure 106: ***Fifty Shades of Grey*, James.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.96 *The Stone Sky, Jemisin*

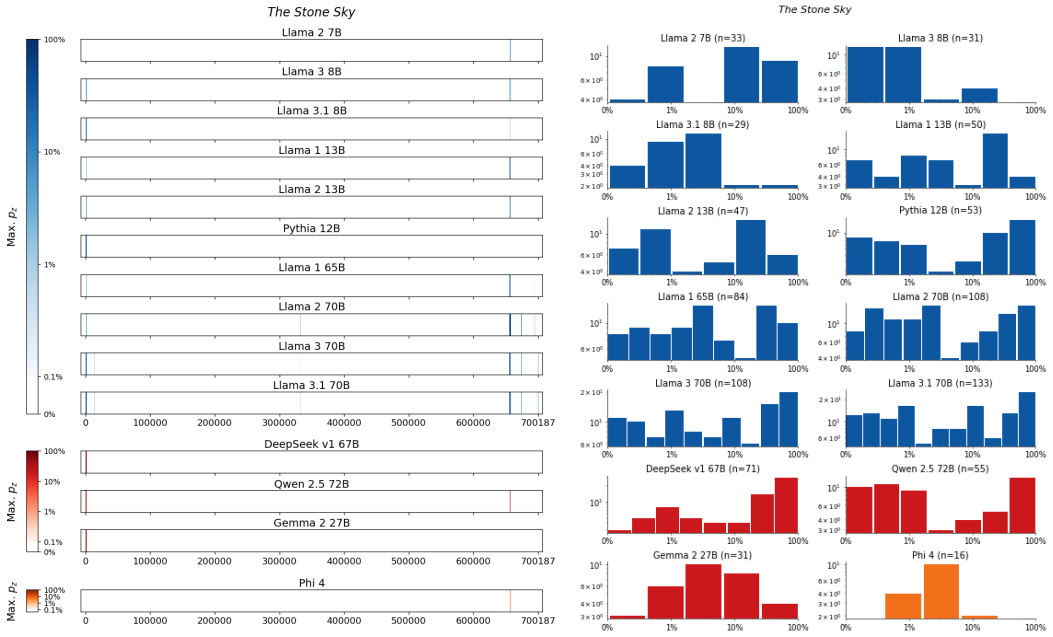


Figure 107: ***The Stone Sky*, Jemisin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.97 *Ulysses, Joyce*

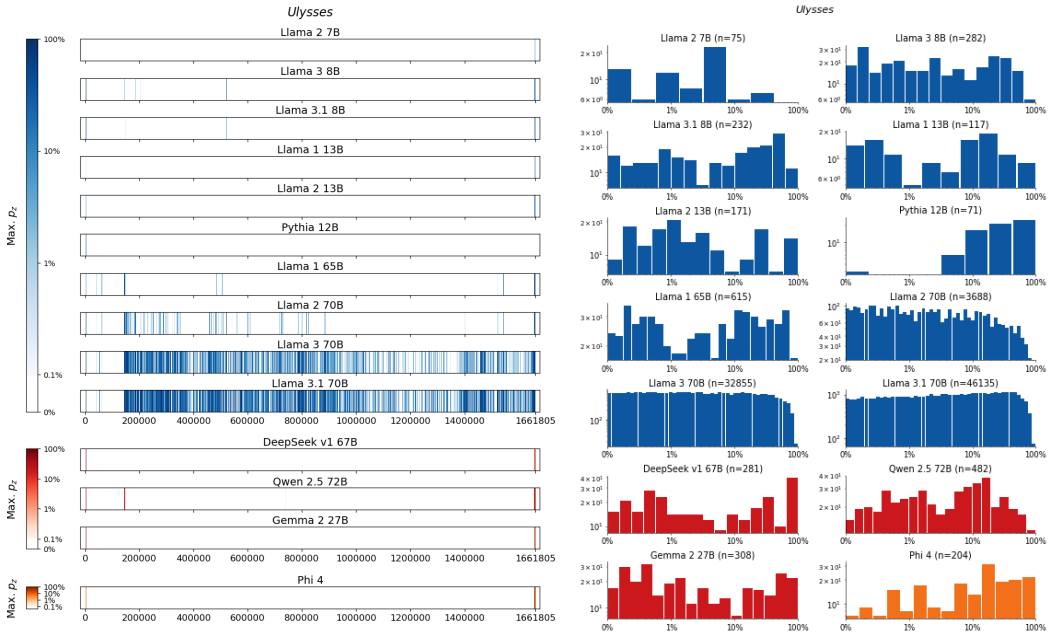


Figure 108: ***Ulysses*, Joyce.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$). This is a public domain book, so PHI 4 is not a negative control.

E.3.98 *Sandman Slim, Kadrey*

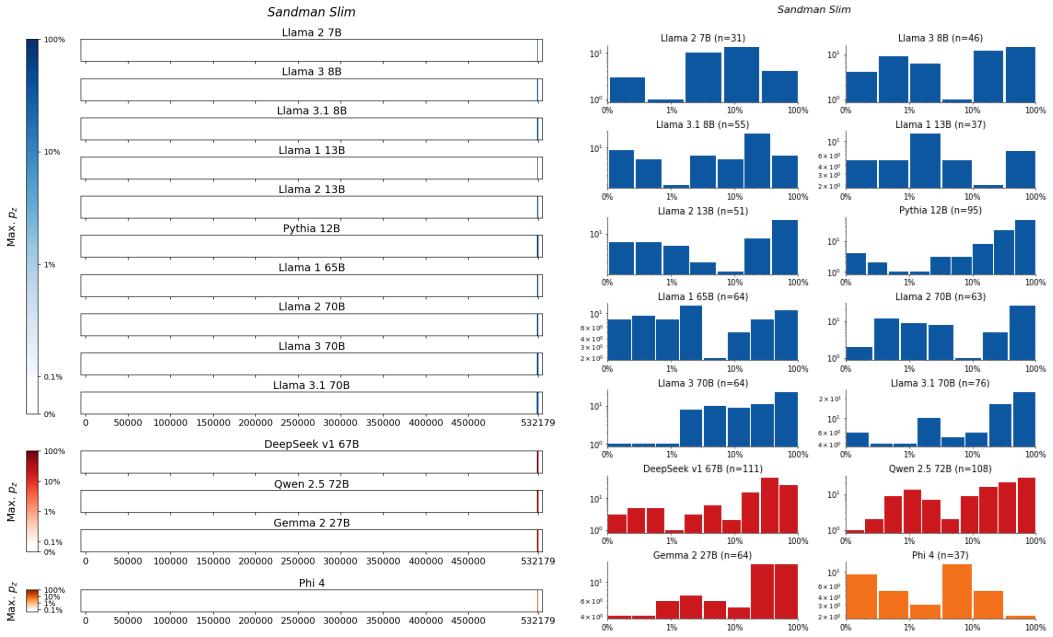


Figure 109: ***Sandman Slim*, Kadrey.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.99 *Ethnography after Antiquity,* Kaldellis

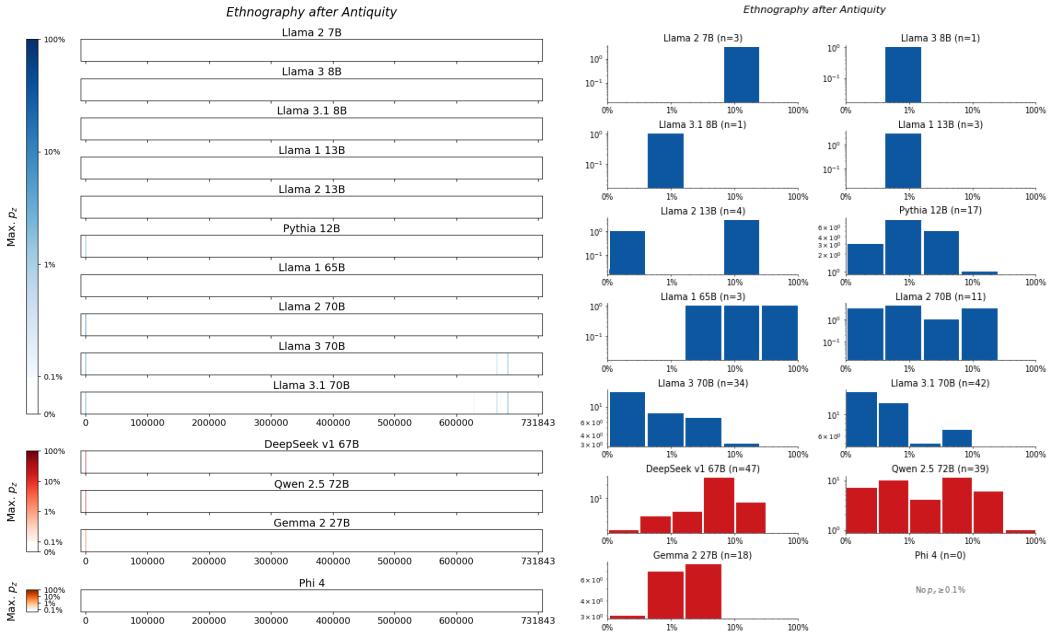


Figure 110: ***Ethnography after Antiquity,* Kaldellis.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\min} = 0.1\%$).

E.3.100 *Who Is Rich?,* Klam

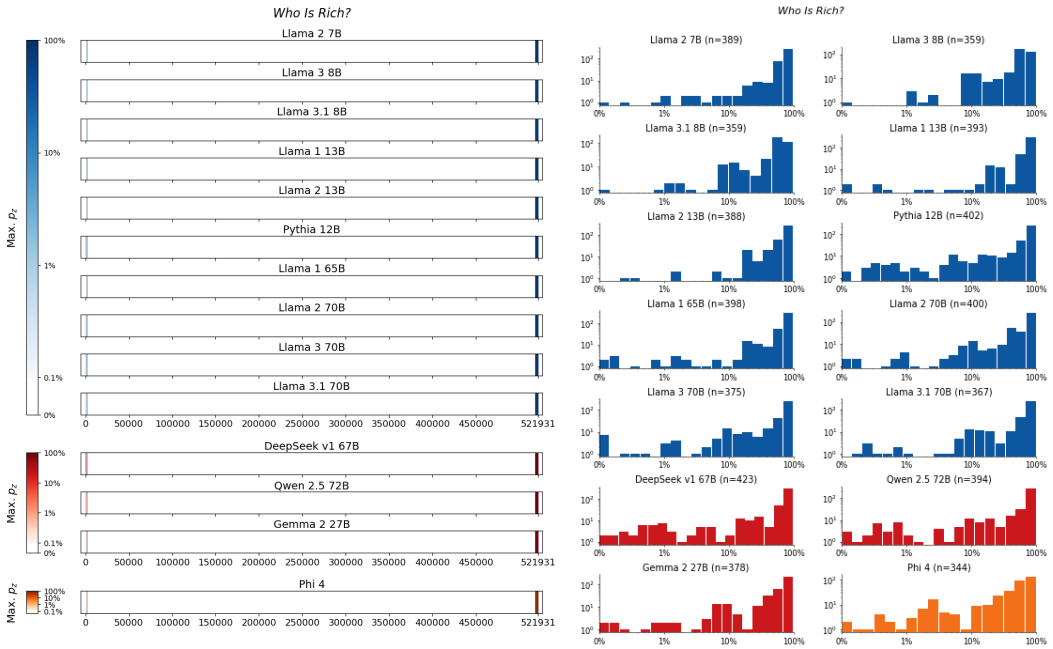


Figure 111: ***Who Is Rich?,* Klam.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\min} = 0.1\%$).

E.3.101 *The Servants of Twilight,* Koontz

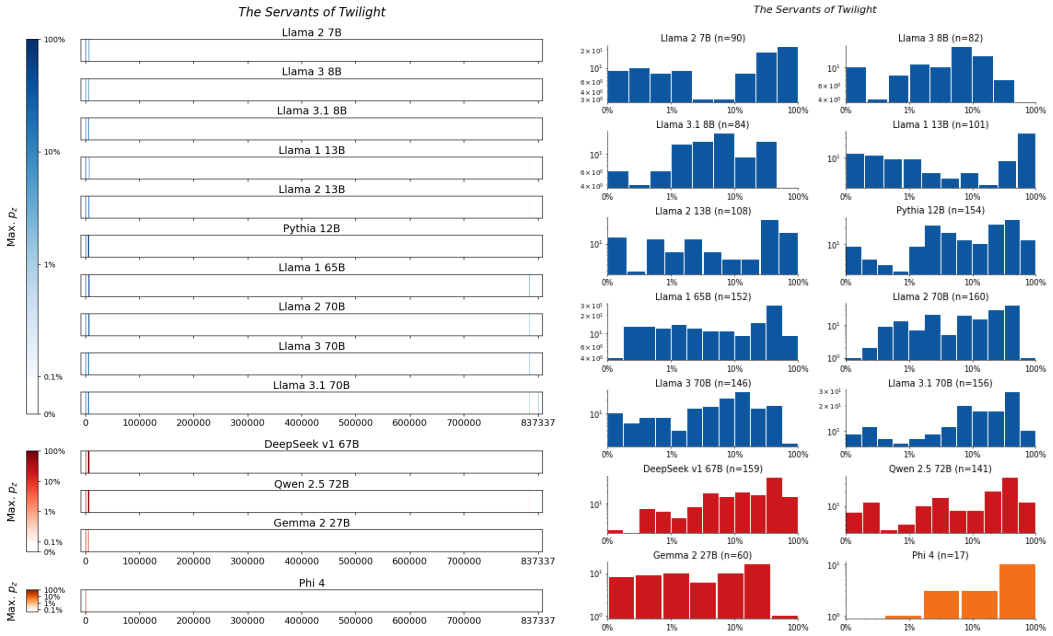


Figure 112: ***The Servants of Twilight,* Koontz.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.102 *Tai Chi for Depression,* Kuhn

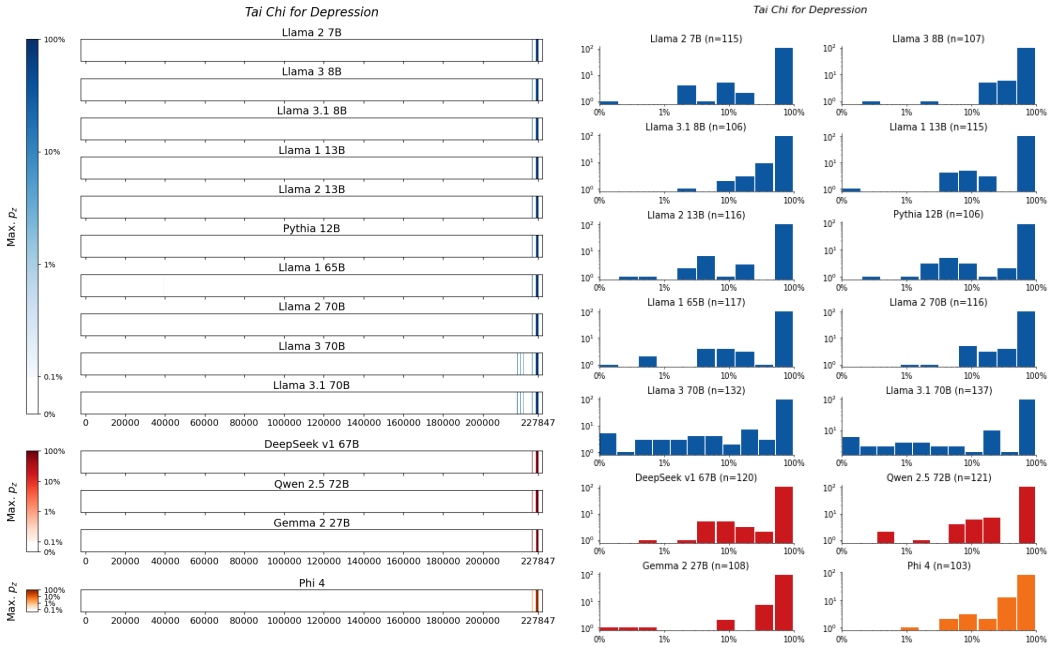


Figure 113: ***Tai Chi for Depression,* Kuhn.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.103 *The Tide Was Always High,* ***Kun***

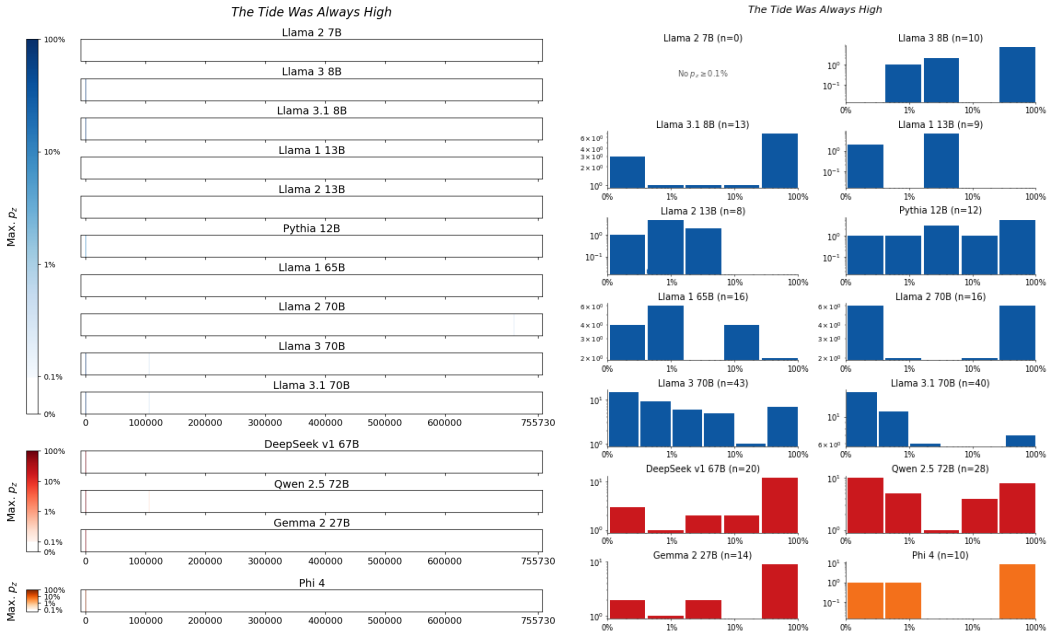


Figure 114: ***The Tide Was Always High,* Kun.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.104 *Girl in Translation,* ***Kwok***

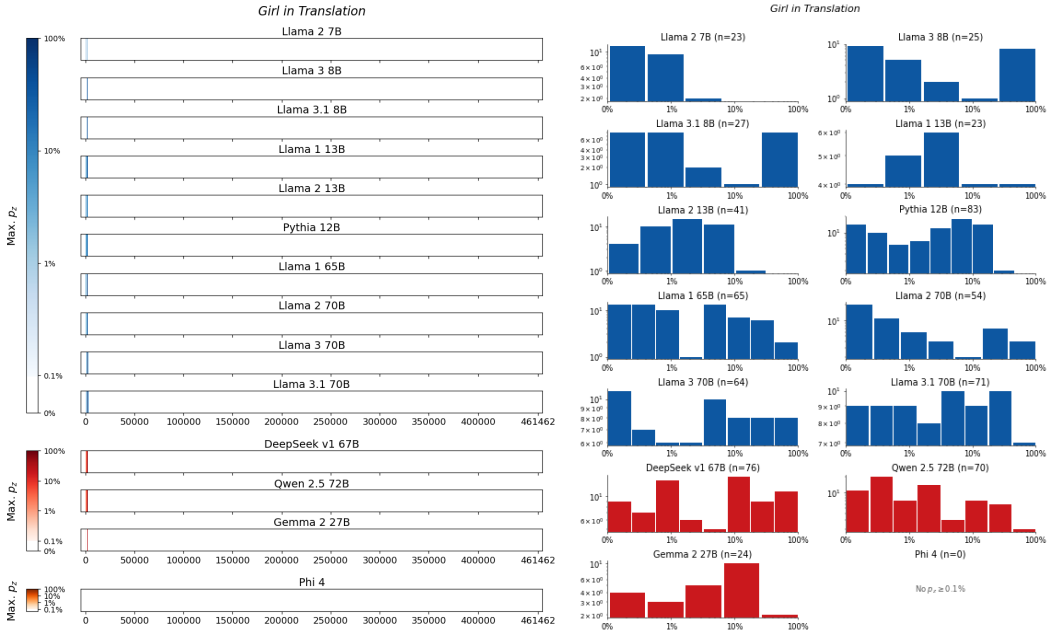


Figure 115: ***Girl in Translation,* Kwok.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.105 A Wrinkle in Time, L'Engle

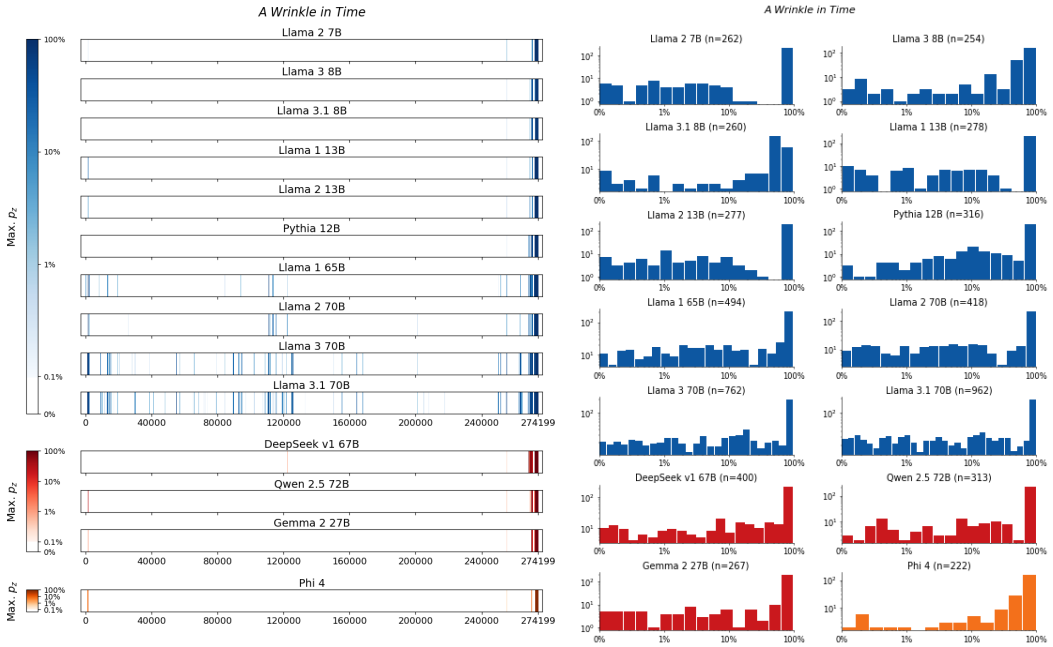


Figure 116: ***A Wrinkle in Time*, L'Engle.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.106 Call Me Brooklyn, Lago

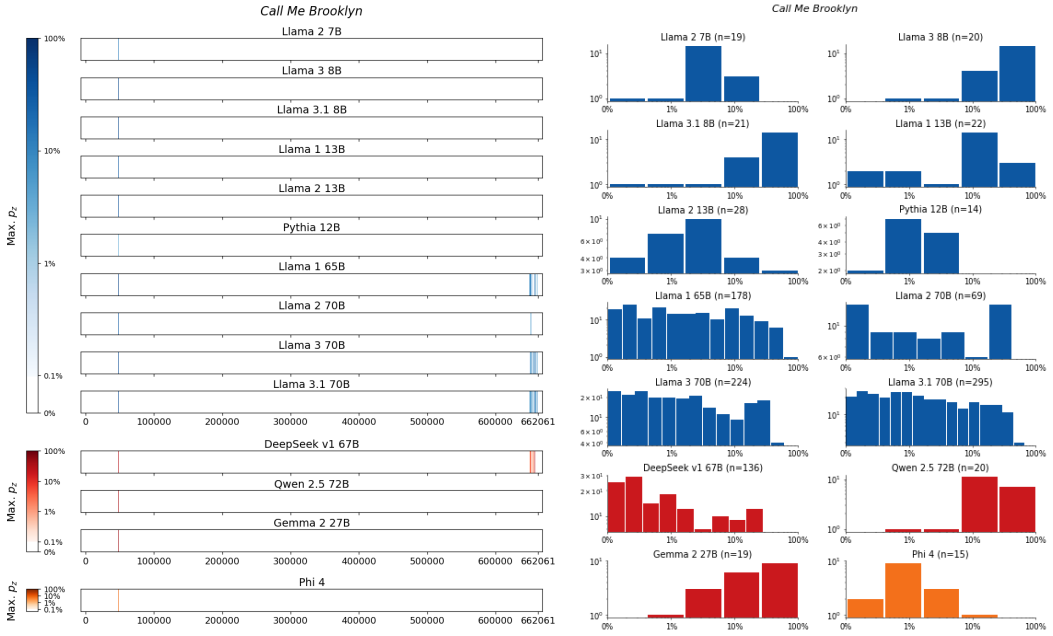


Figure 117: ***Call Me Brooklyn*, Lago.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.107 Dead Wake, Larson

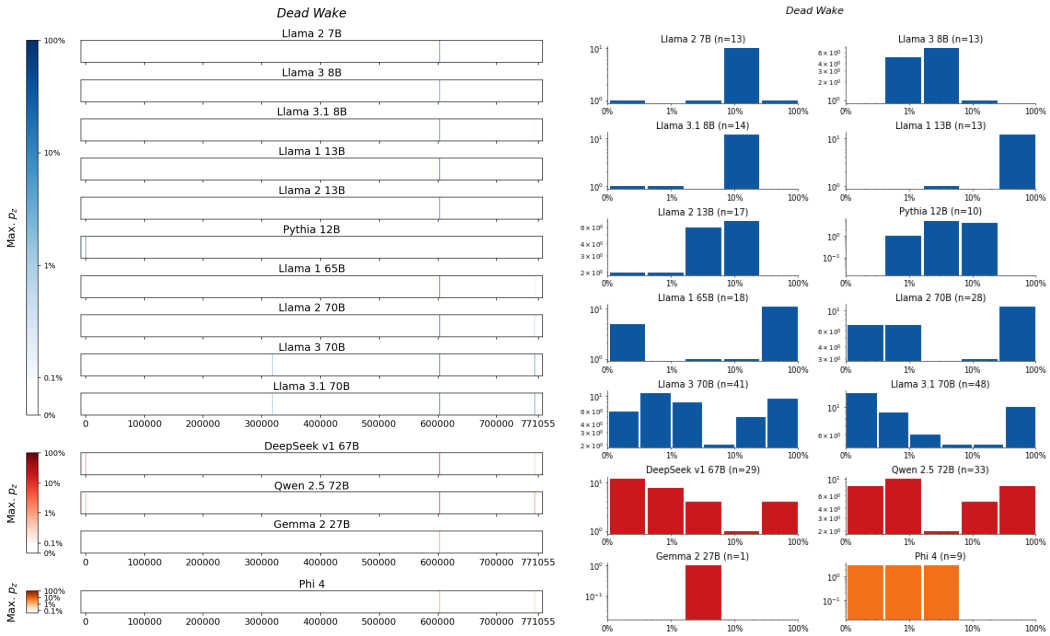


Figure 118: ***Dead Wake*, Larson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.108 The Girl with the Dragon Tattoo, Larsson

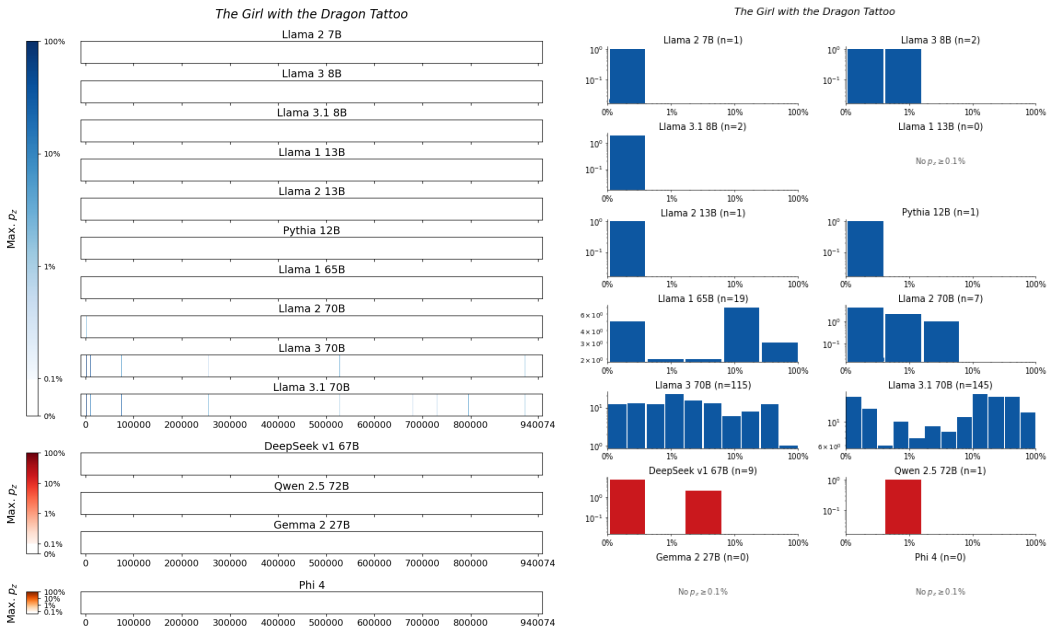


Figure 119: ***The Girl with the Dragon Tattoo*, Larsson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.109 *The Daughter of Odren, Le Guin*

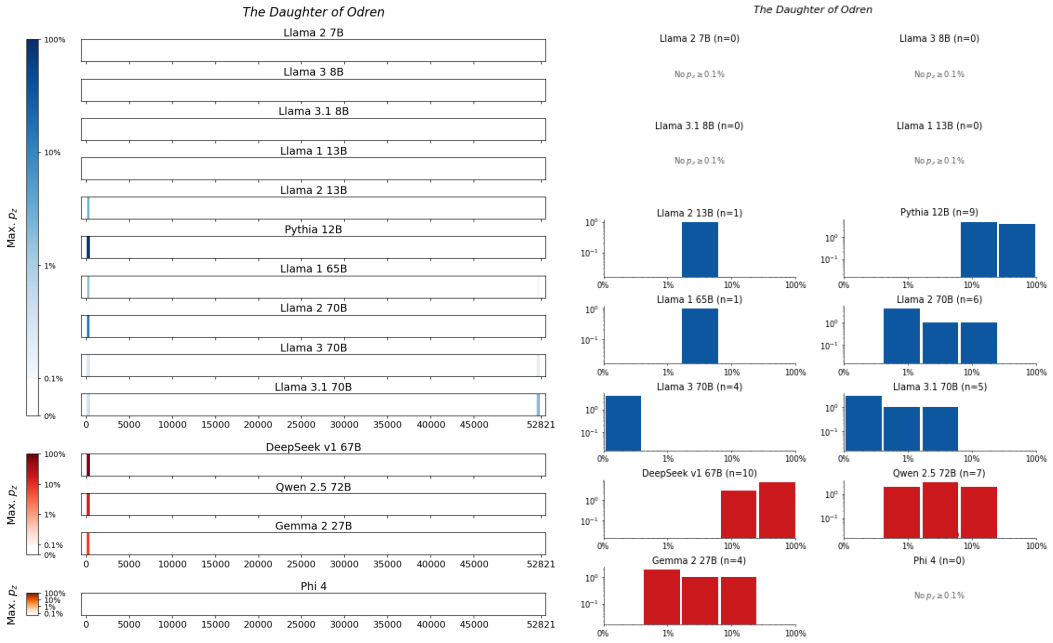


Figure 120: ***The Daughter of Odren*, Le Guin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.110 *The Chronicles of Narnia, Lewis*

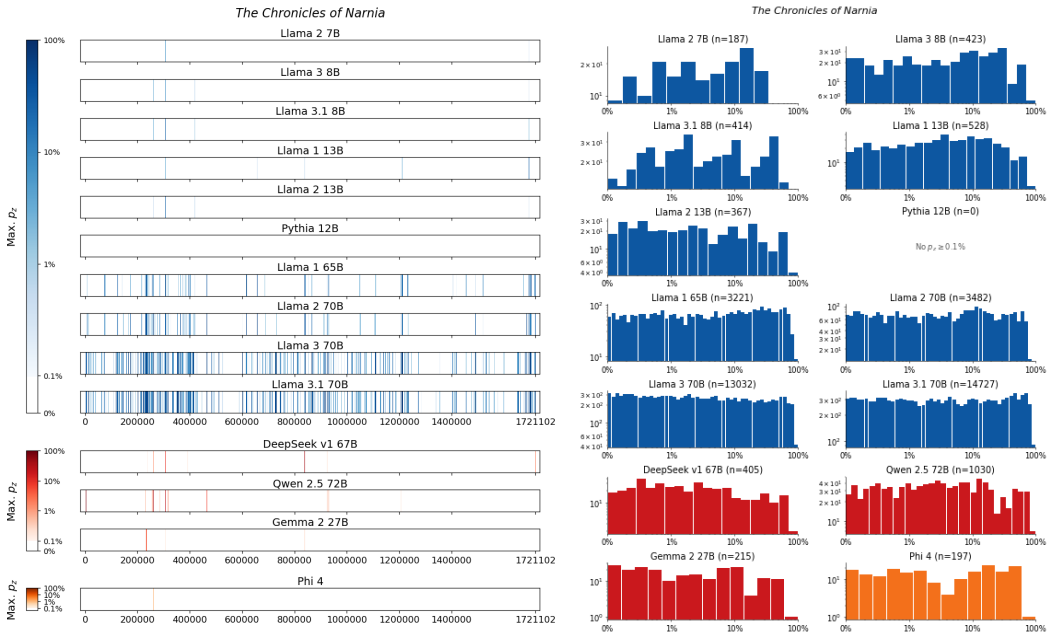


Figure 121: ***The Chronicles of Narnia*, Lewis.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.111 *After I'm Gone,* ***Lippman***

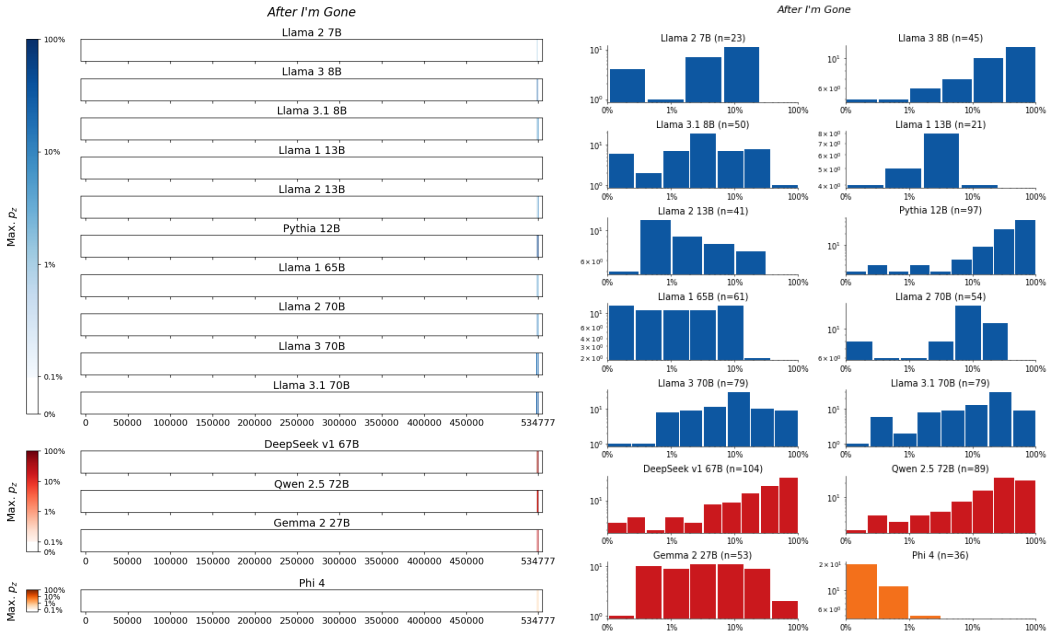


Figure 122: ***After I'm Gone,* Lippman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.112 *Sunburn,* ***Lippman***

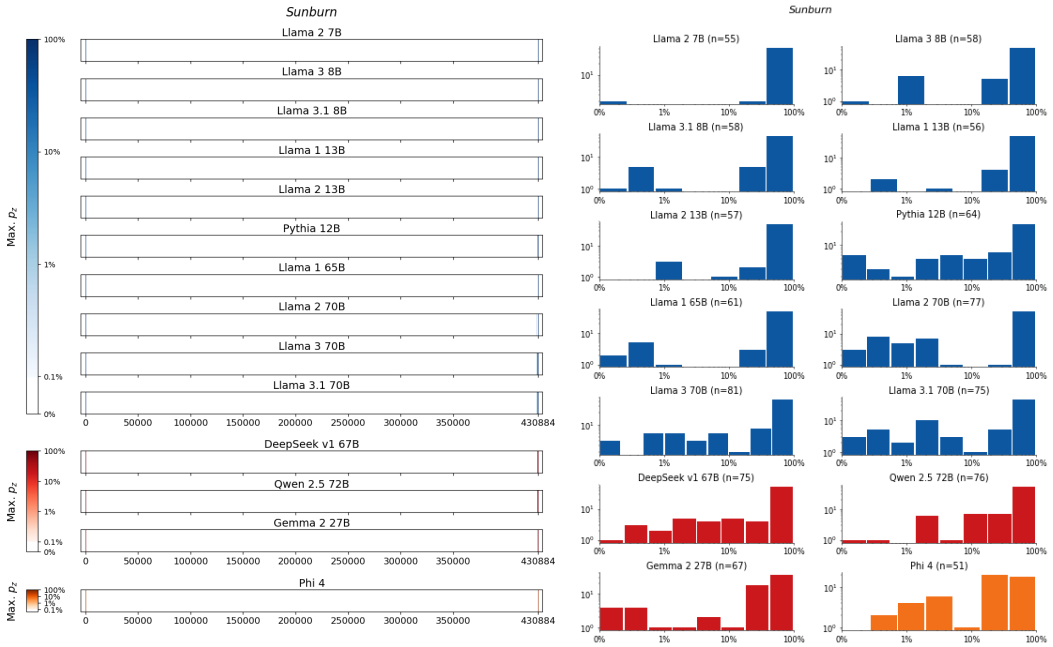


Figure 123: ***Sunburn,* Lippman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.113 Marvel's Spider-Man: Hostile Takeover, Liss

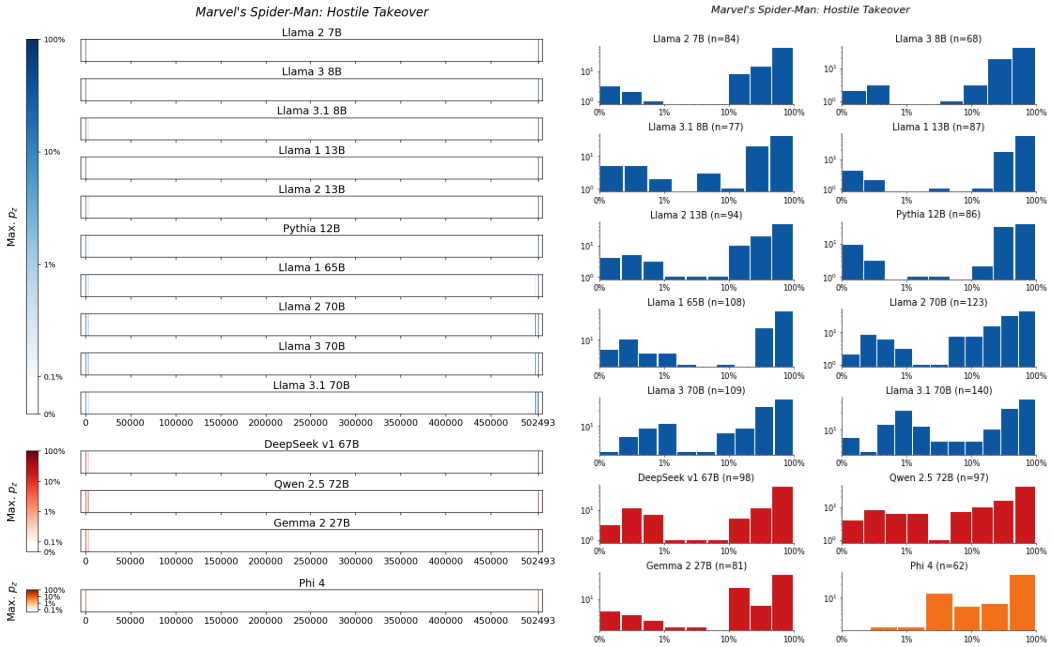


Figure 124: ***Marvel's Spider-Man: Hostile Takeover*, Liss.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.114 Anastasia on Her Own, Lowry

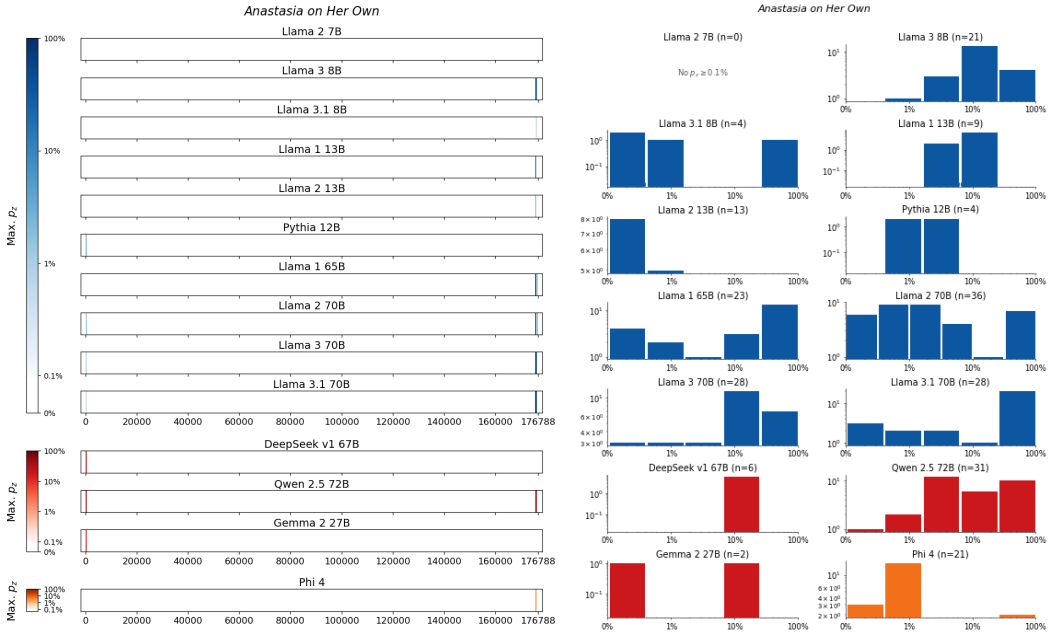


Figure 125: ***Anastasia on Her Own*, Lowry.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.115 Nixon in China, MacMillan

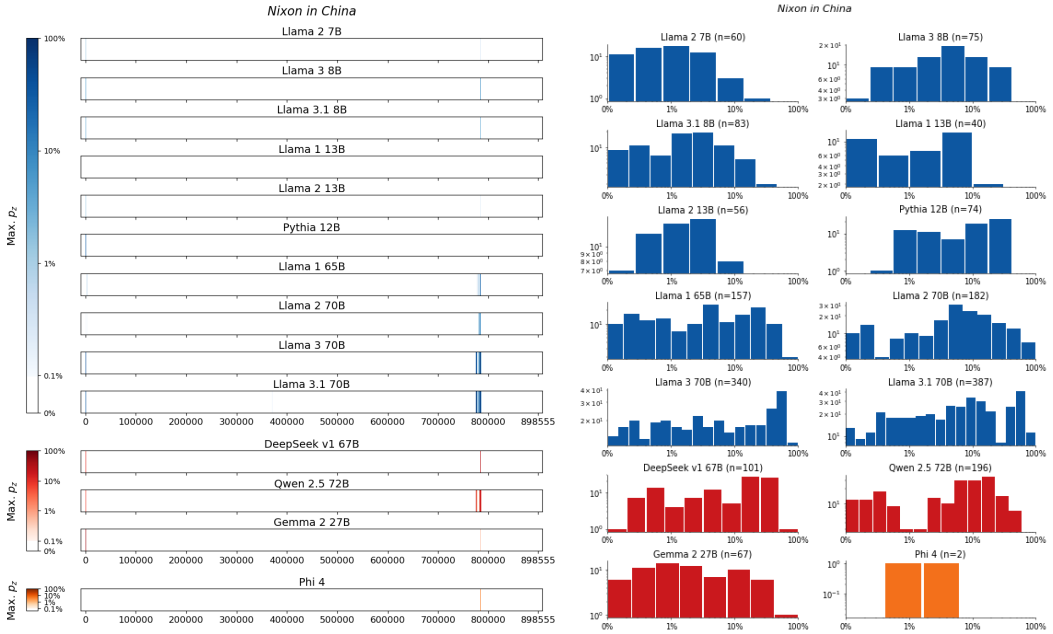


Figure 126: ***Nixon in China*, MacMillan.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.116 The Doomsday Prophecy, Mariani

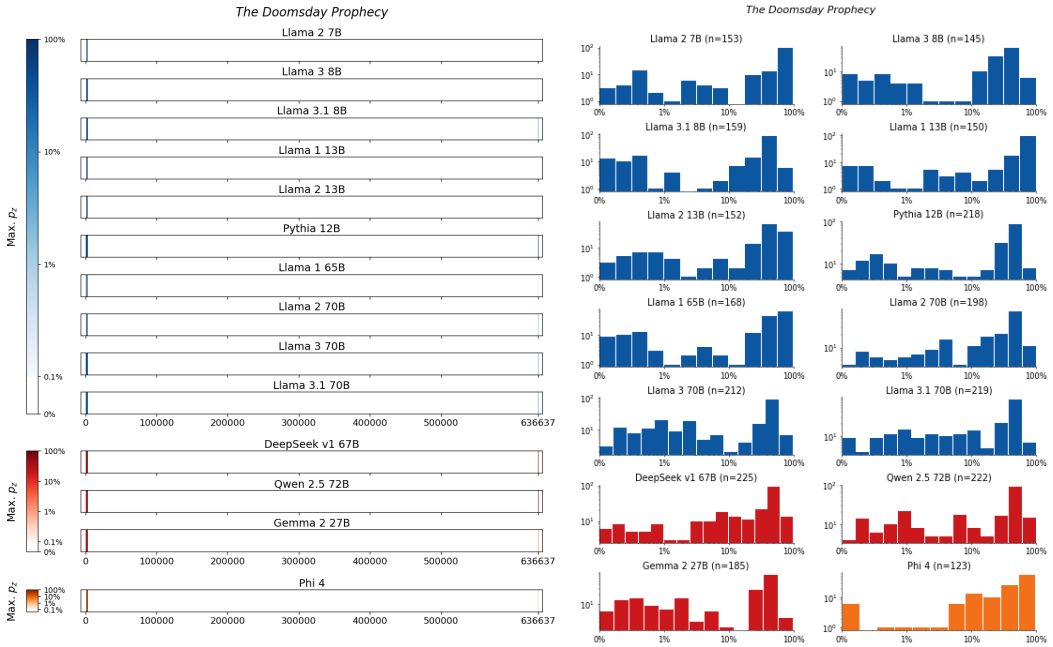


Figure 127: ***The Doomsday Prophecy*, Mariani.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.117 A Game of Thrones, Martin

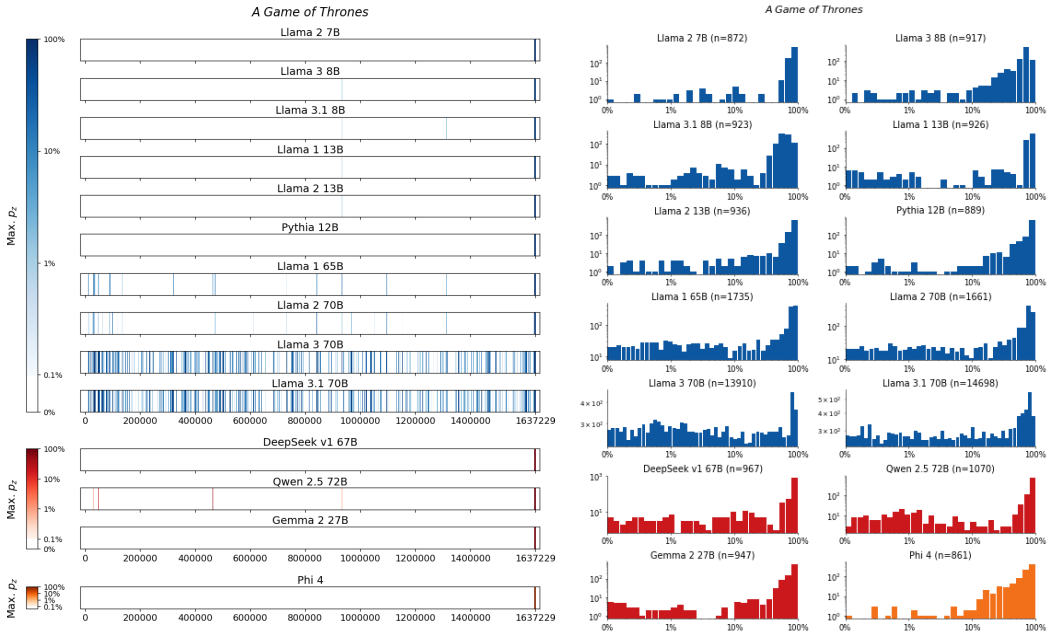


Figure 128: ***A Game of Thrones*, Martin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.118 Rough-Hewn Land, Meldahl

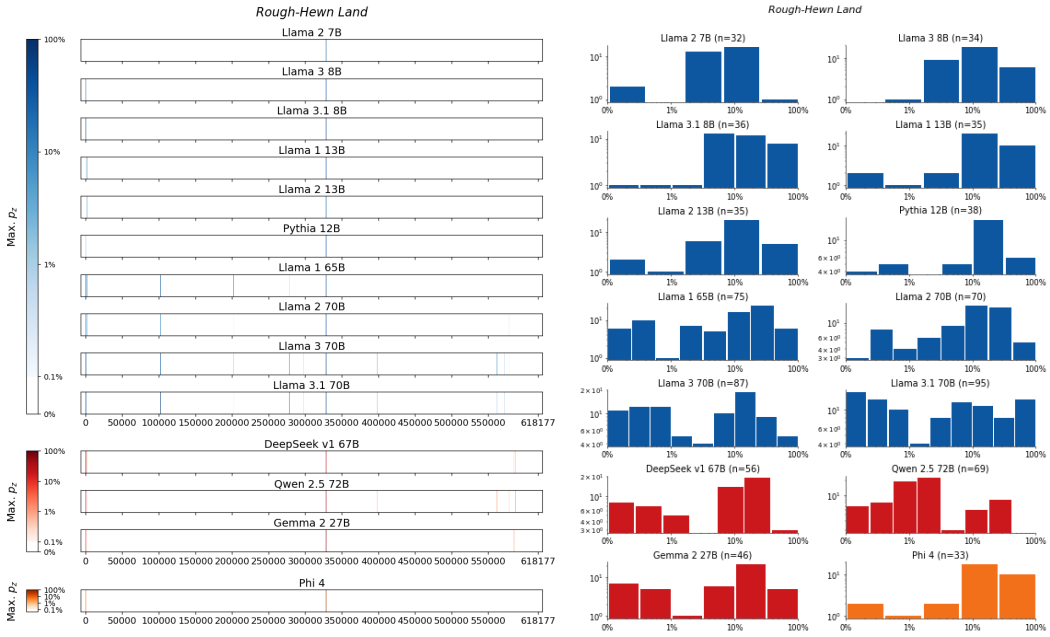


Figure 129: ***Rough-Hewn Land*, Meldahl.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.119 Twilight, Meyer

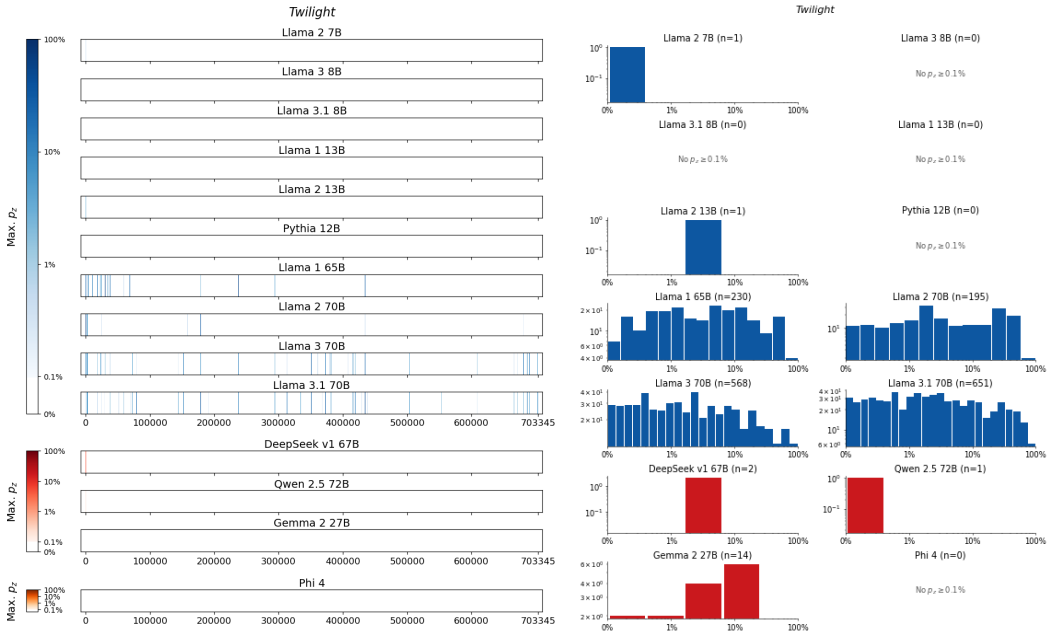


Figure 130: ***Twilight*, Meyer.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.120 The Duchess War, Milan

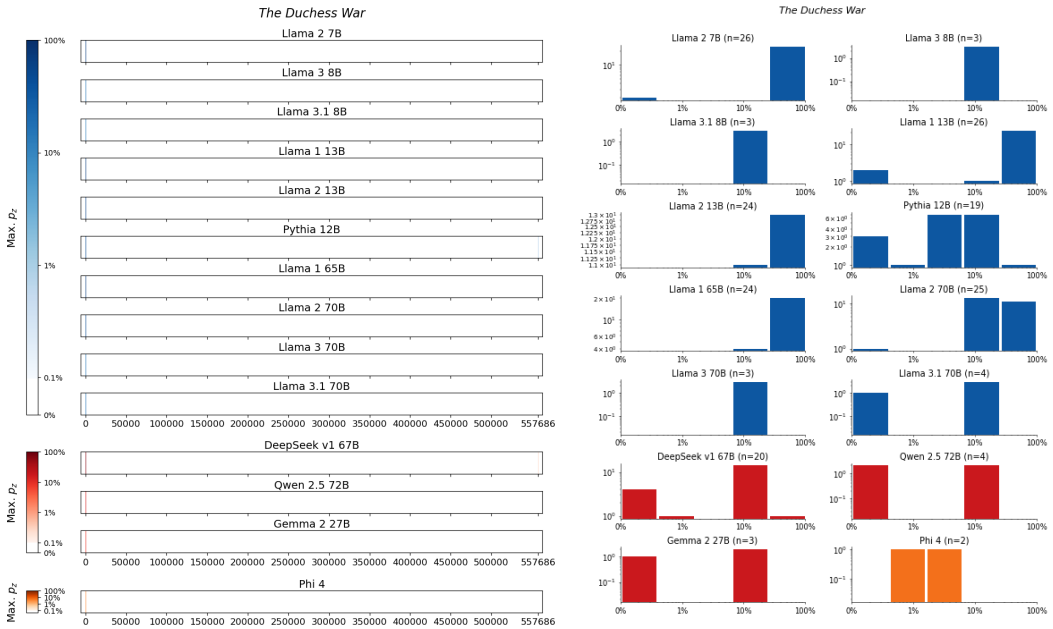


Figure 131: ***The Duchess War*, Milan.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.121 *On Liberty, Mill*

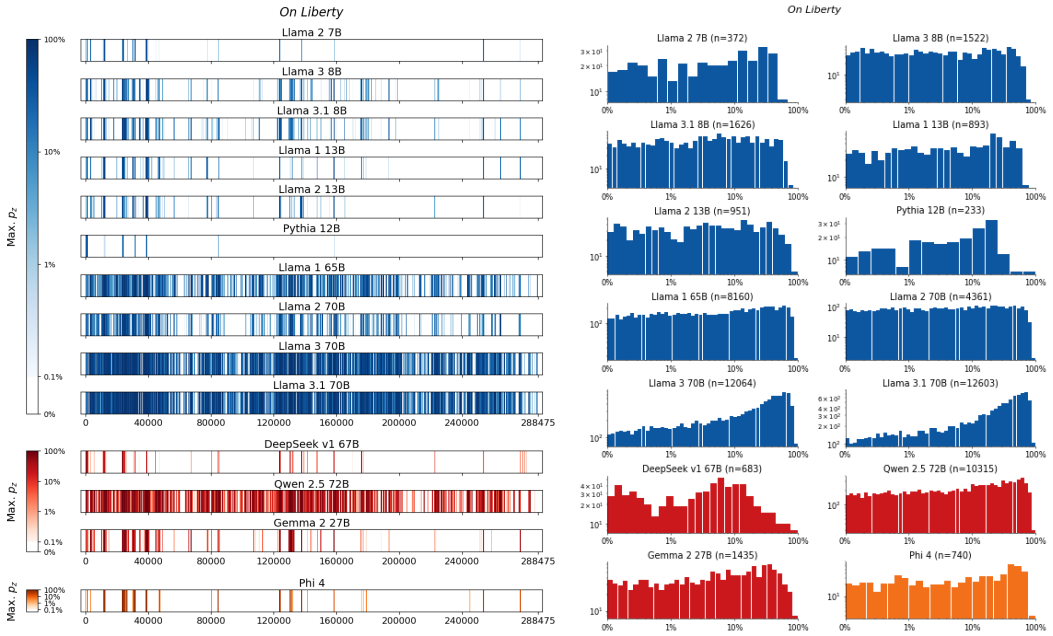


Figure 132: ***On Liberty*, Mill.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.122 *Coal Creek, Miller*

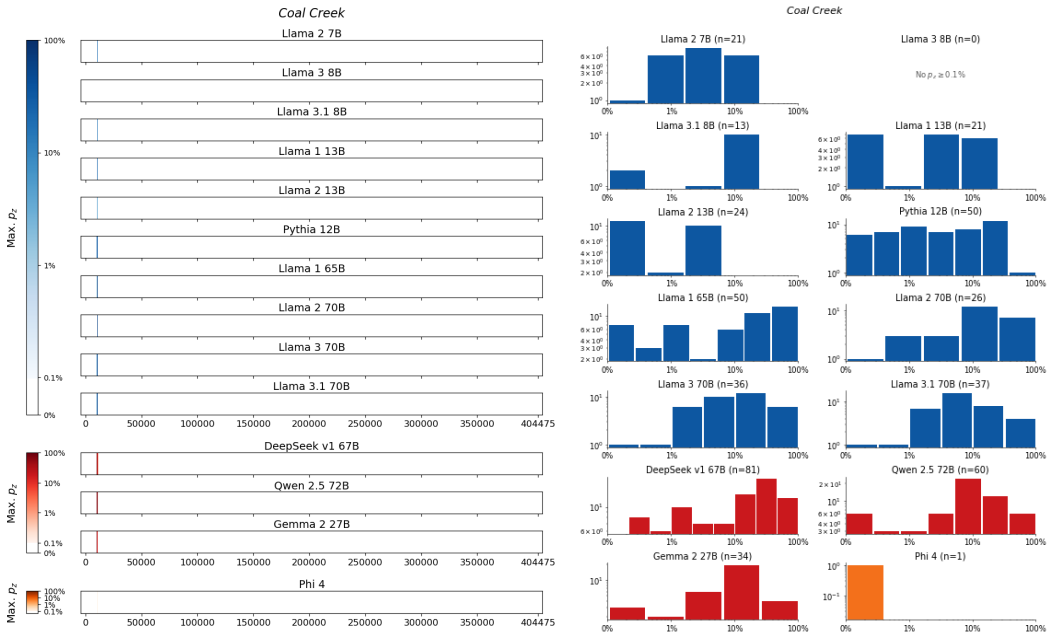


Figure 133: ***Coal Creek*, Miller.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.123 *Winnie the Pooh, Milne*

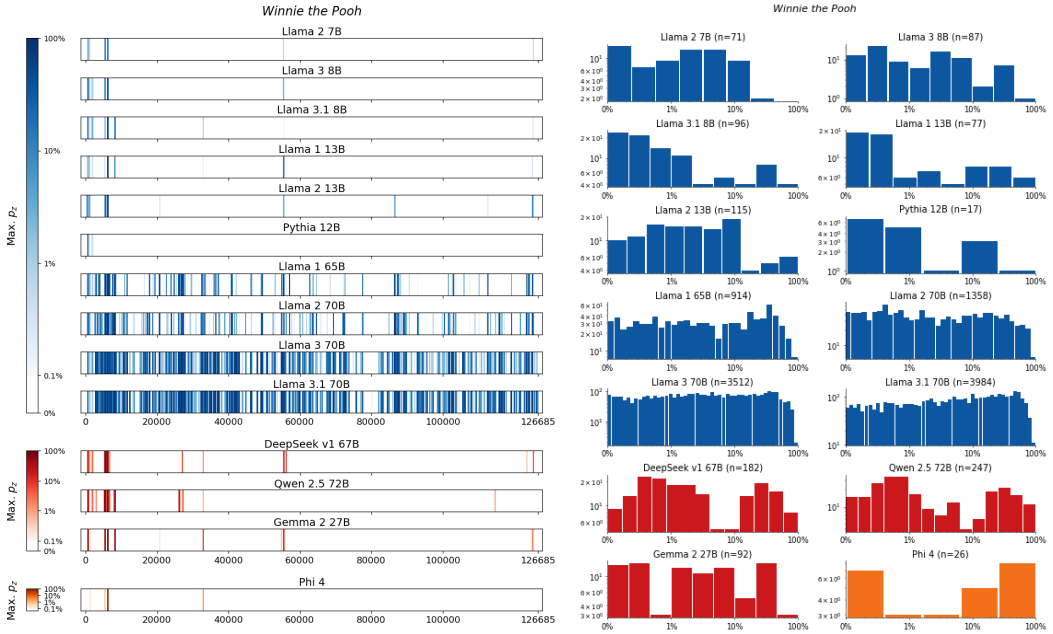


Figure 134: ***Winnie the Pooh*, Milne.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.124 *Catching the Sky, Moore & O'Brien*

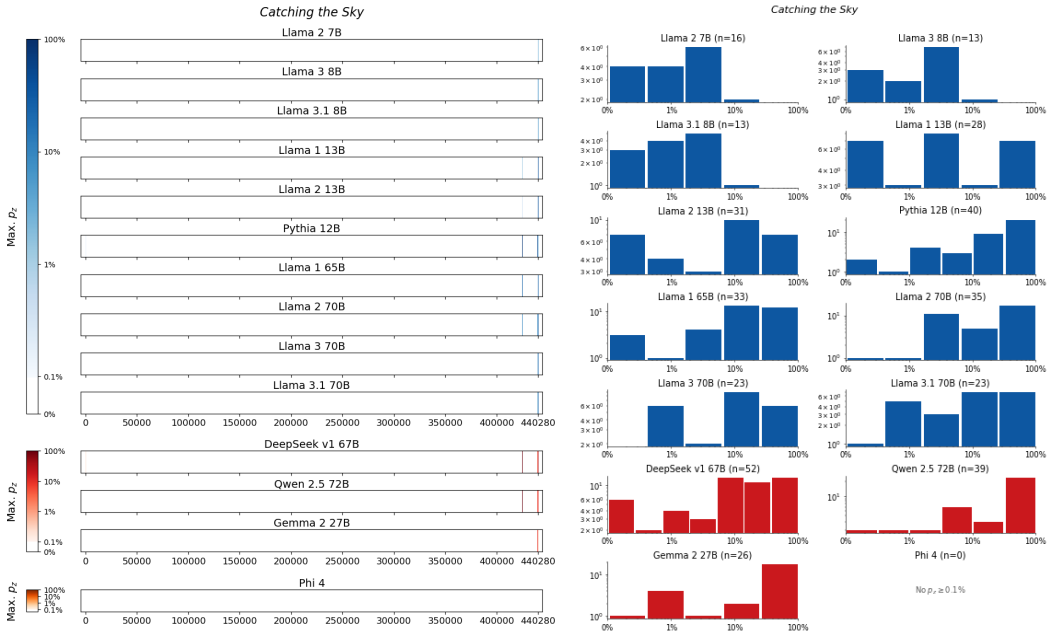


Figure 135: ***Catching the Sky*, Moore & O'Brien.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.125 The Heretic Queen, Moran

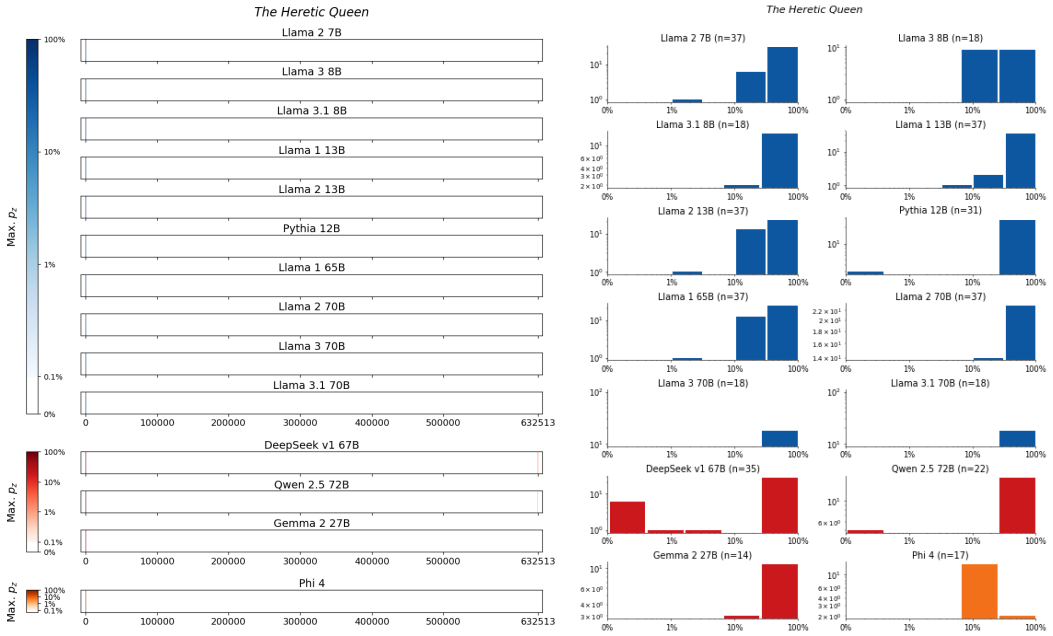


Figure 136: ***The Heretic Queen*, Moran.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.126 The Gondola Maker, Morelli

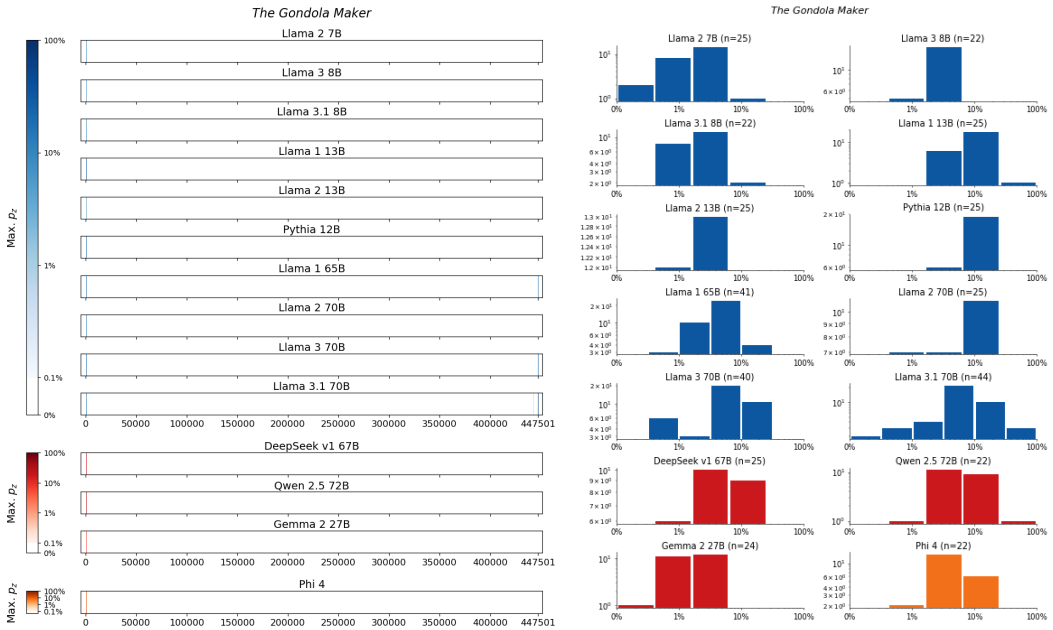


Figure 137: ***The Gondola Maker*, Morelli.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.127 *Songs in Ordinary Time,* Morris

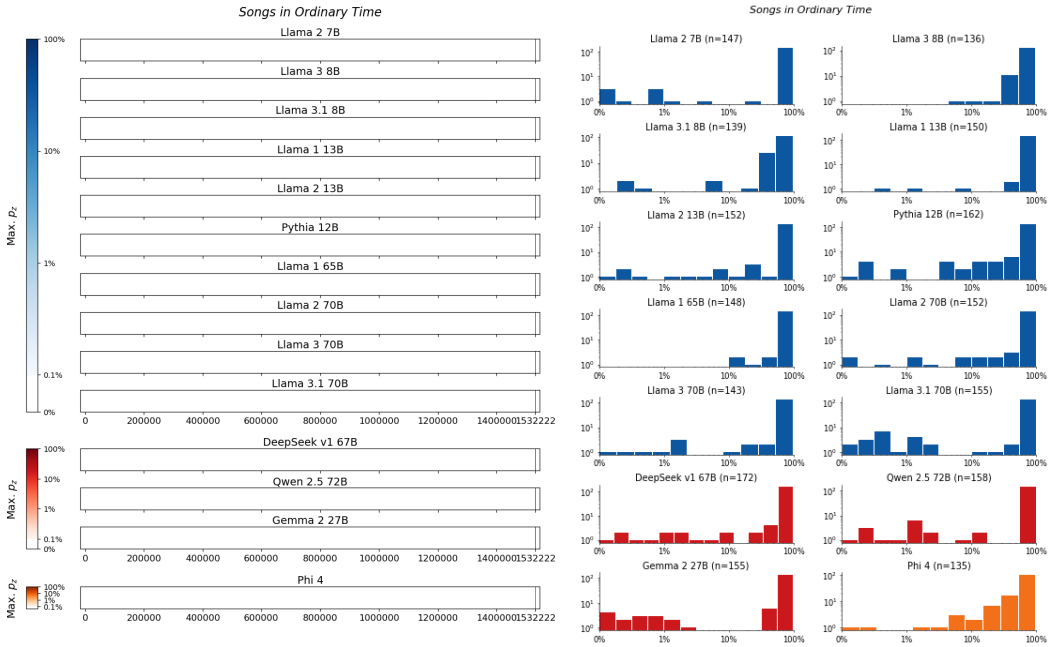


Figure 138: ***Songs in Ordinary Time,* Morris.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.128 *Beloved,* Morrison

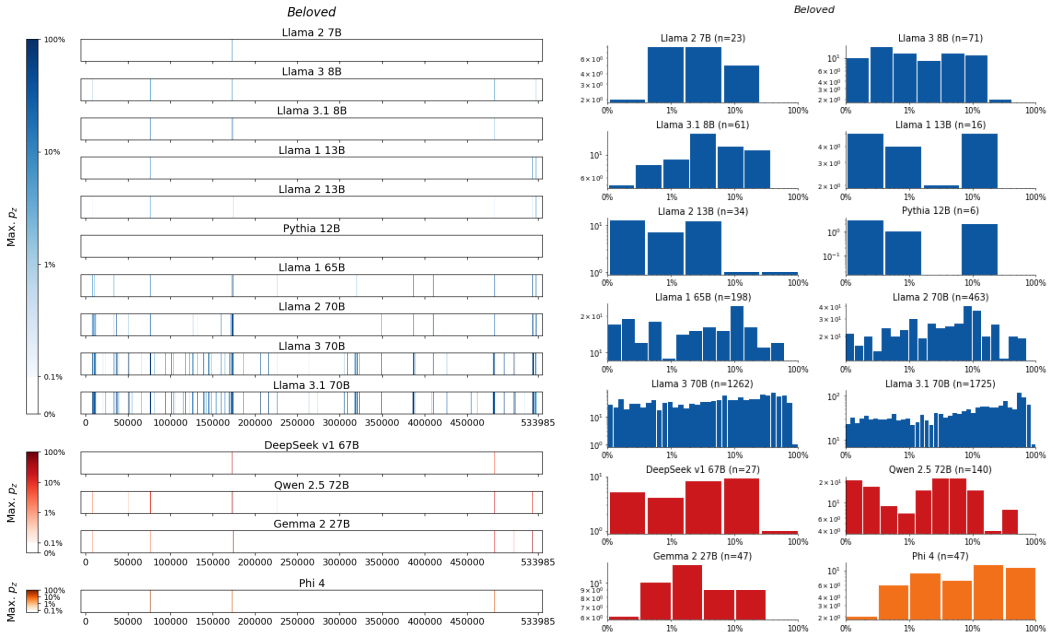


Figure 139: ***Beloved,* Morrison.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.129 Norwegian Wood, Murakami

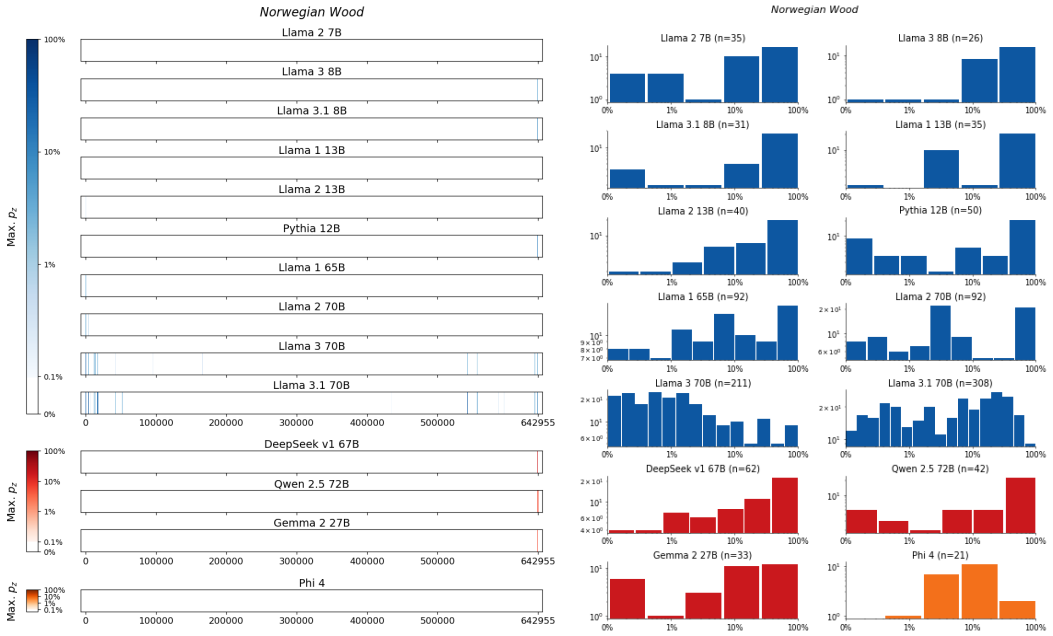


Figure 140: ***Norwegian Wood*, Murakami.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.130 Eat More Plants, Nielsen

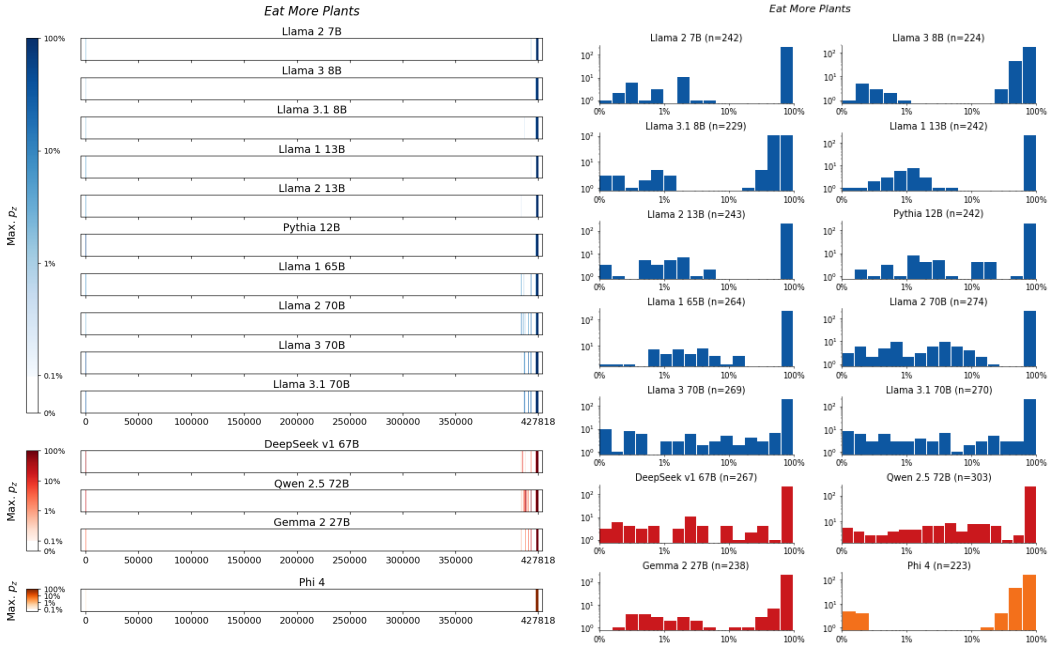


Figure 141: ***Eat More Plants*, Nielsen.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.131 Polaris, Northrop

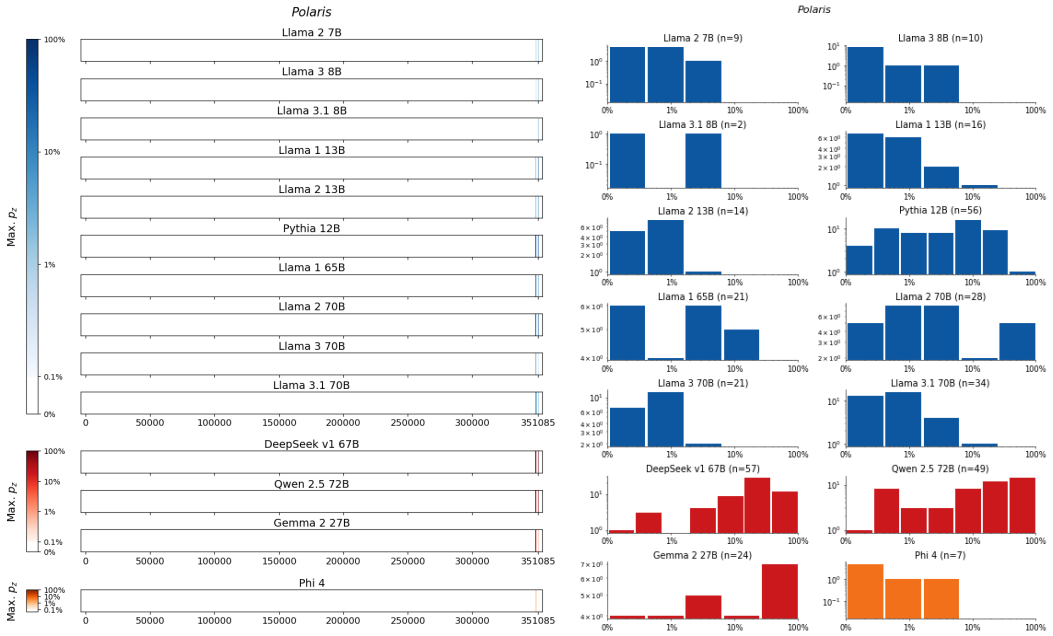


Figure 142: ***Polaris*, Northrop.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.132 Pagans, O'Donnell

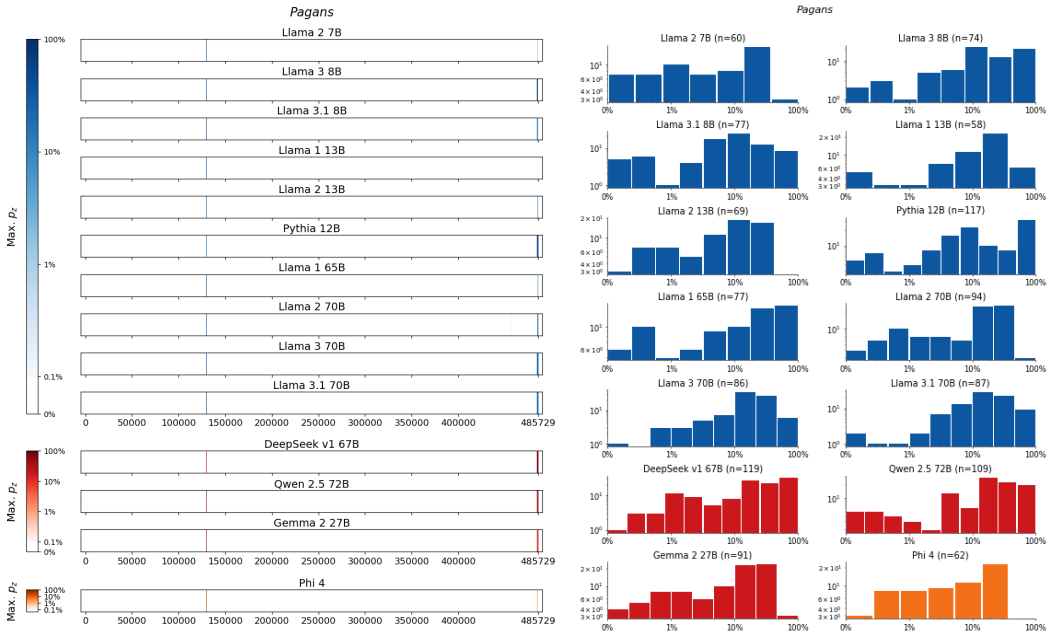


Figure 143: ***Pagans*, O'Donnell.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.133 *Windfall, O'Sullivan*

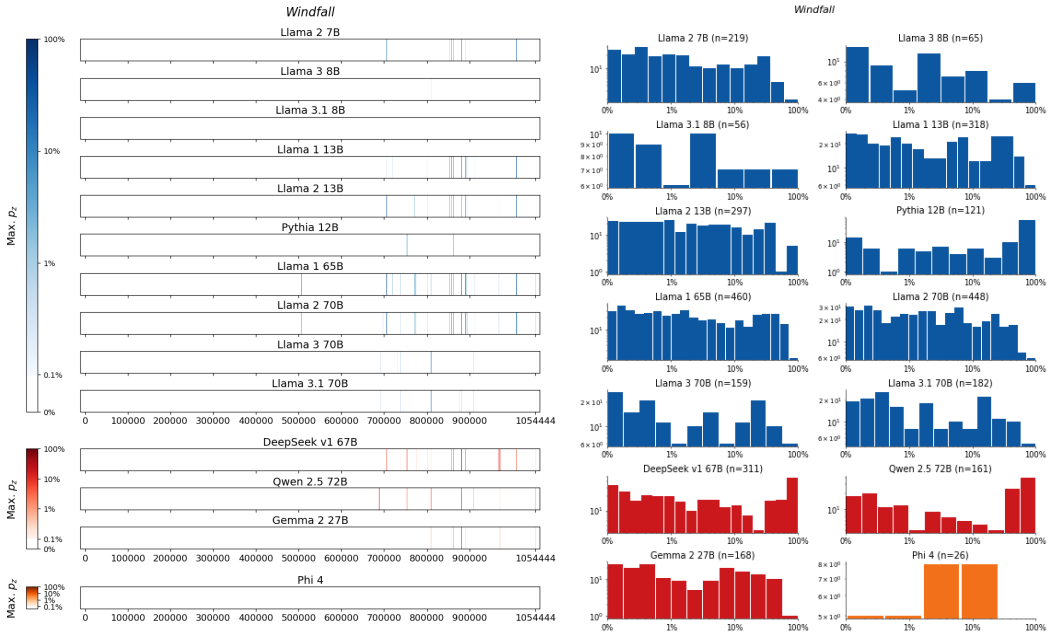


Figure 144: ***Windfall, O'Sullivan.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.134 *The Memory Police, Ogawa*

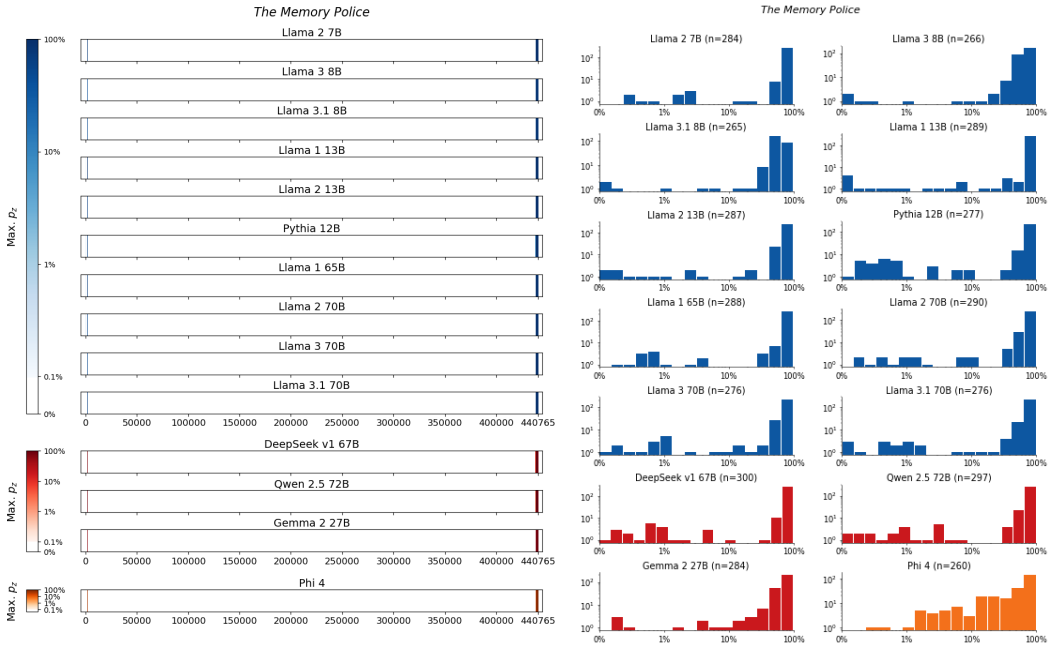


Figure 145: ***The Memory Police, Ogawa.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.135 *Winter Sisters, Oliveira*

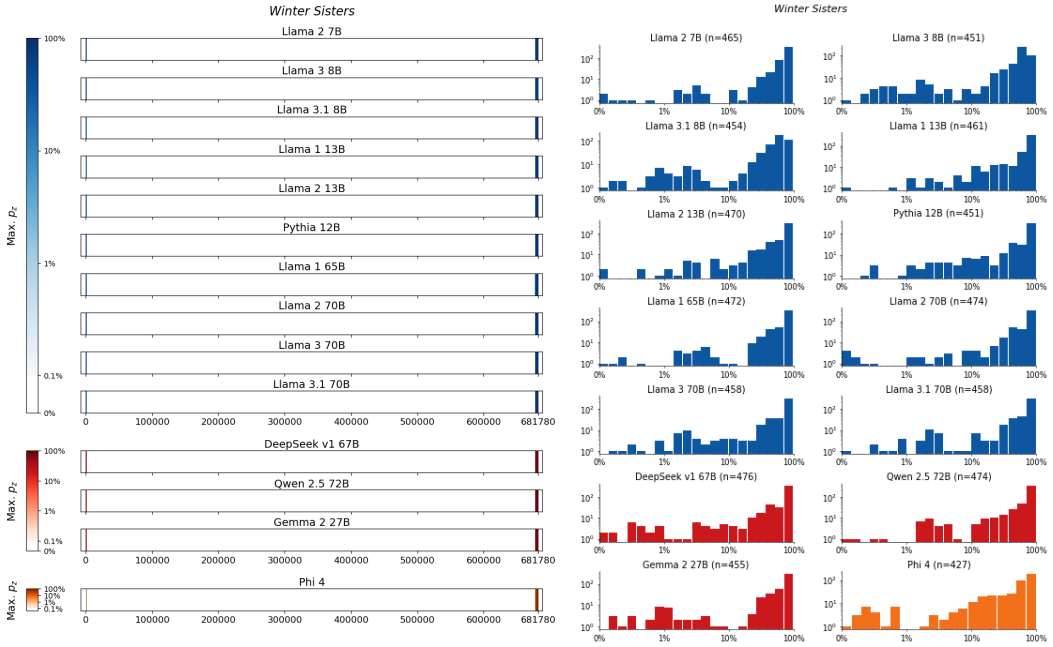


Figure 146: ***Winter Sisters, Oliveira.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.136 *Nineteen Eighty-Four, Orwell*

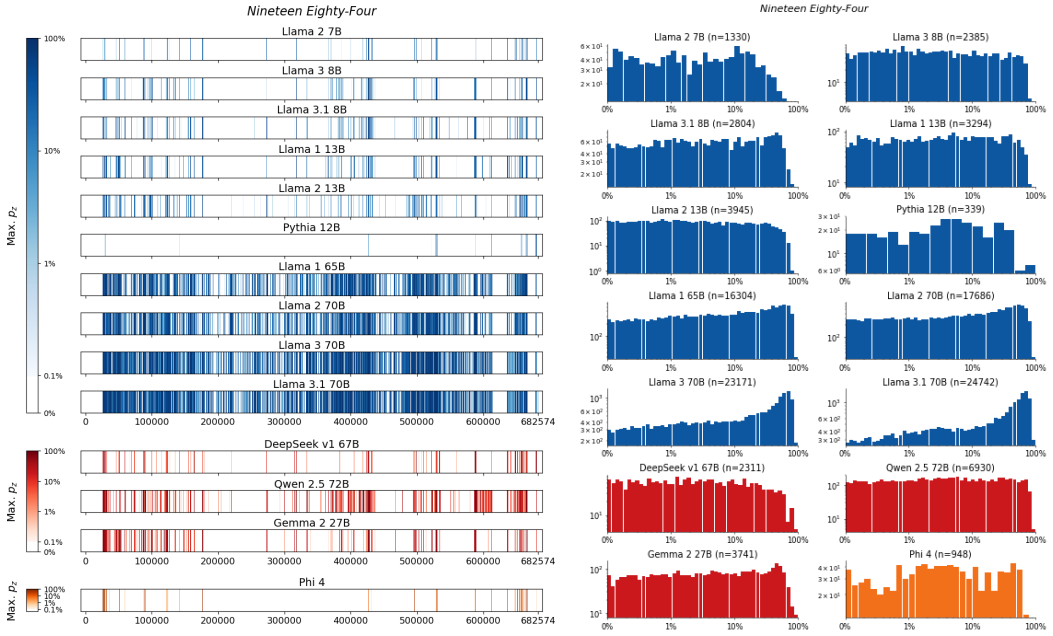


Figure 147: ***Nineteen Eighty-Four, Orwell.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.137 *Fight Club*, Palahniuk

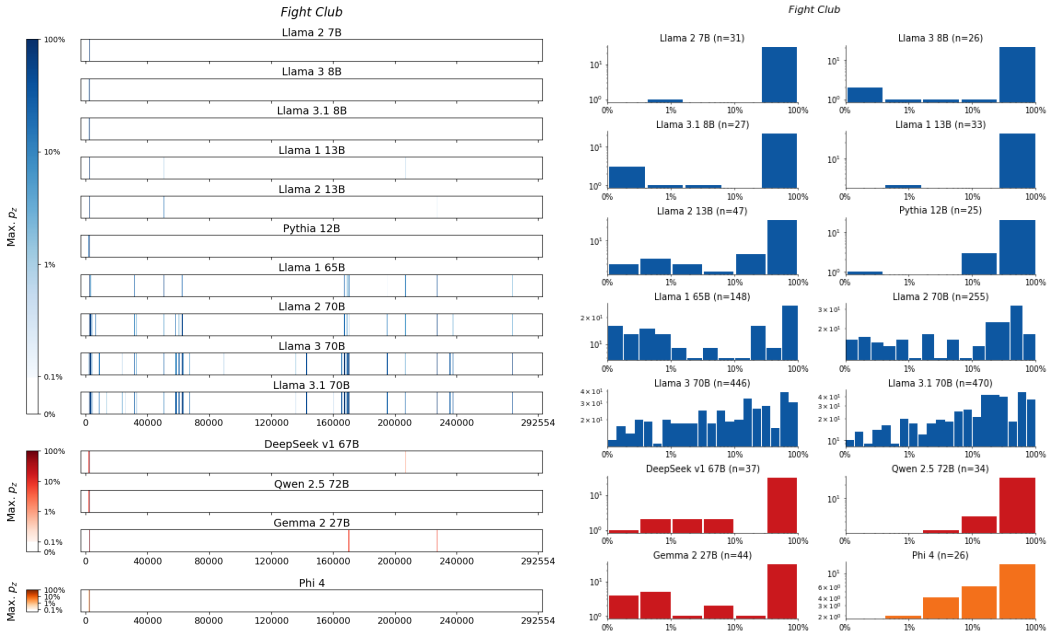


Figure 148: ***Fight Club*, Palahniuk.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.138 *The Complete Joy of Homebrewing*, Papazian

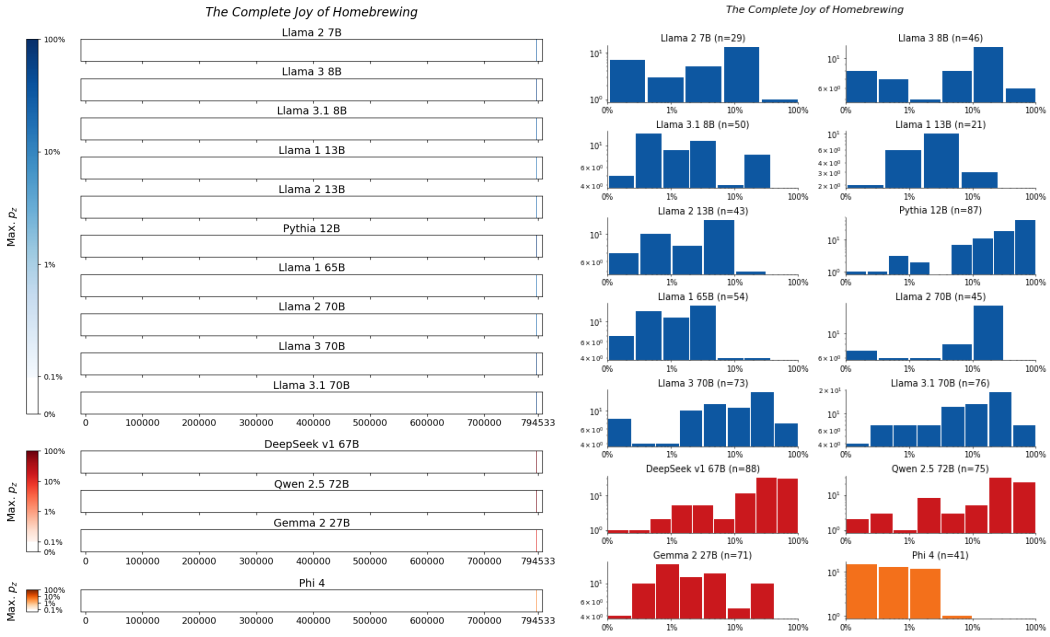


Figure 149: ***The Complete Joy of Homebrewing*, Papazian.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.139 *The Cult of Loving Kindness*, *Park*

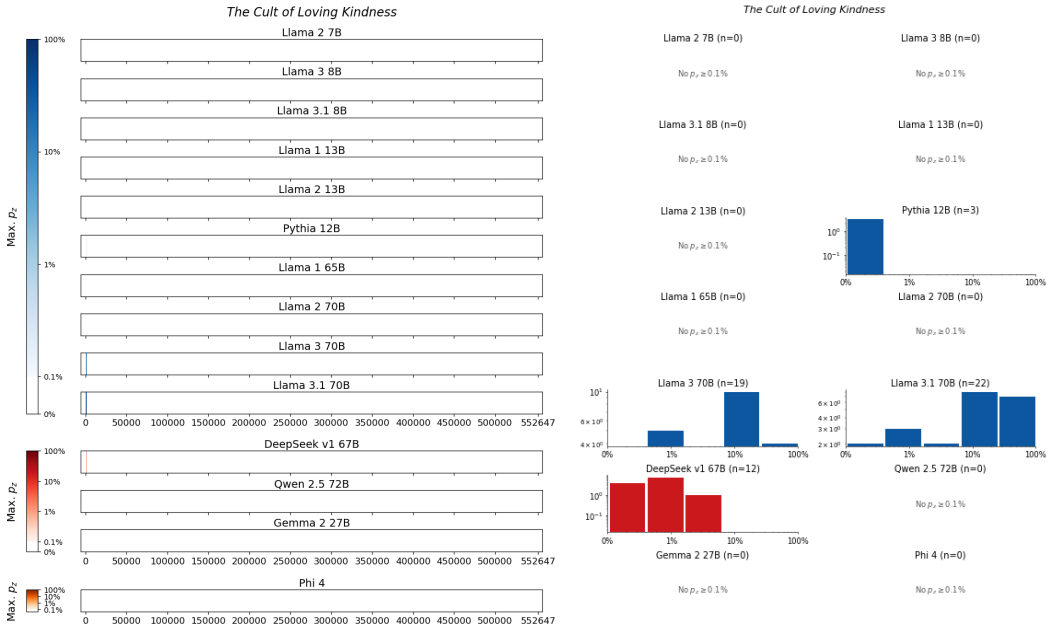


Figure 150: ***The Cult of Loving Kindness*, Park.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.140 *Along Came a Spider*, *Patterson*

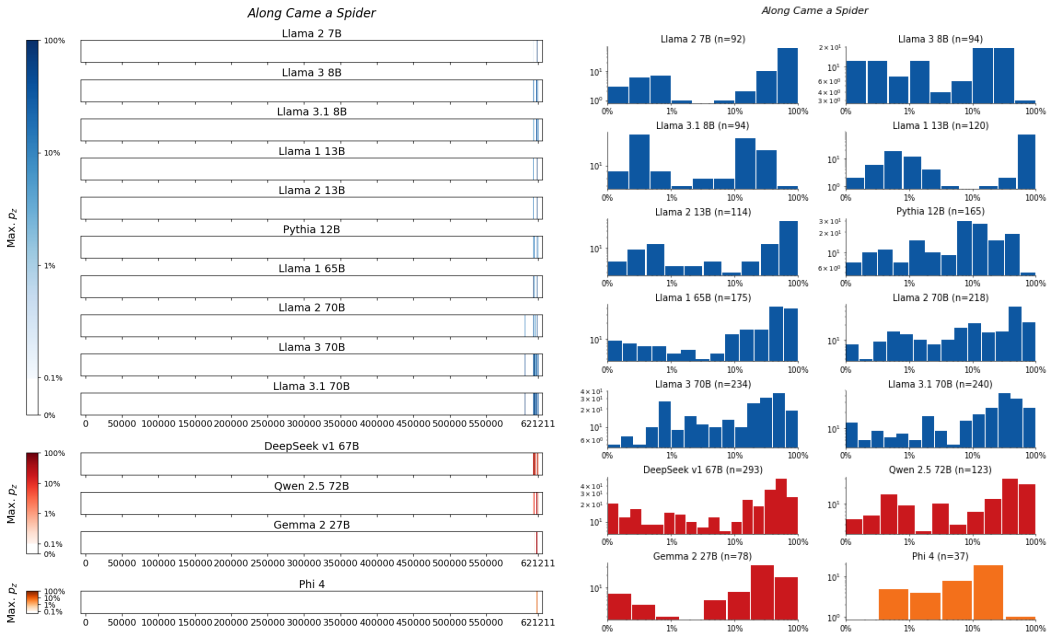


Figure 151: ***Along Came a Spider*, Patterson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.141 *Payard Cookies, Payard & McBride*

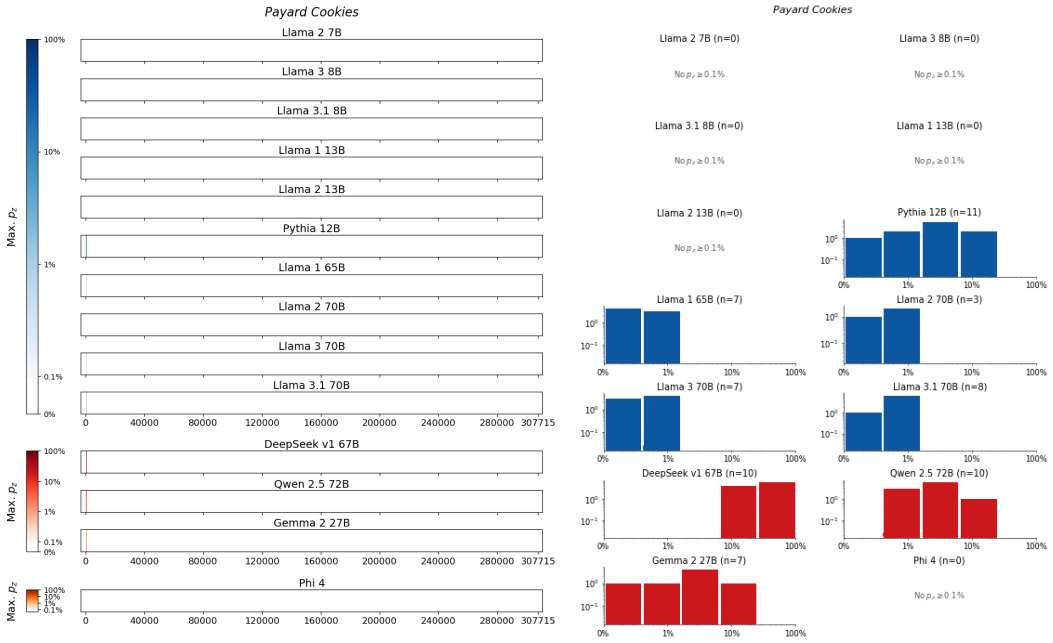


Figure 152: ***Payard Cookies*, Payard & McBride.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.142 *Essential Pepin Desserts, Pépin*

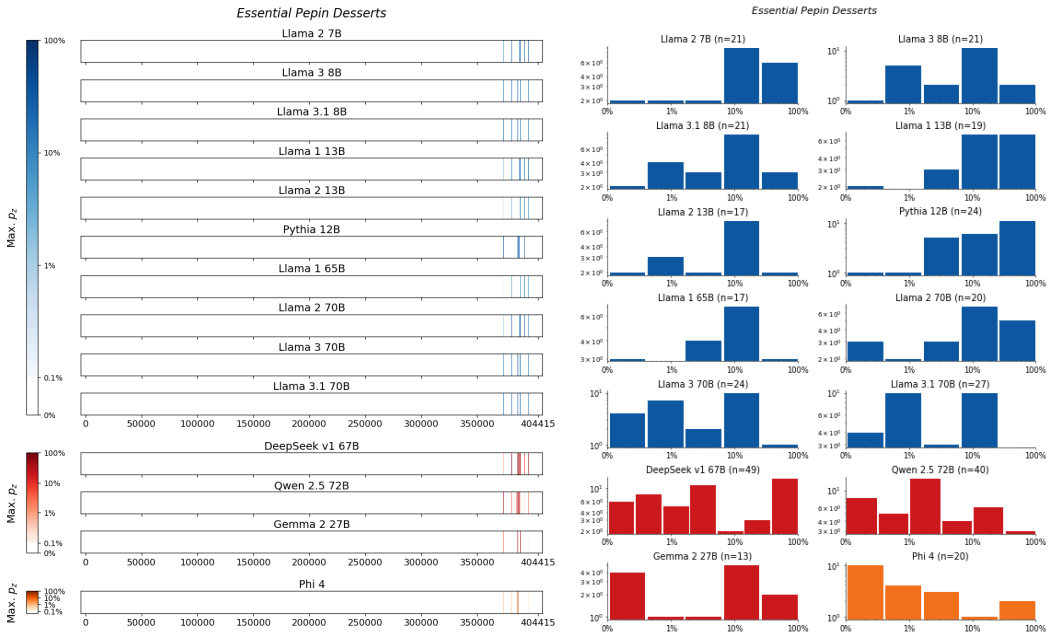


Figure 153: ***Essential Pepin Desserts*, Pépin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.143 *Why New Orleans Matters,* Piazza

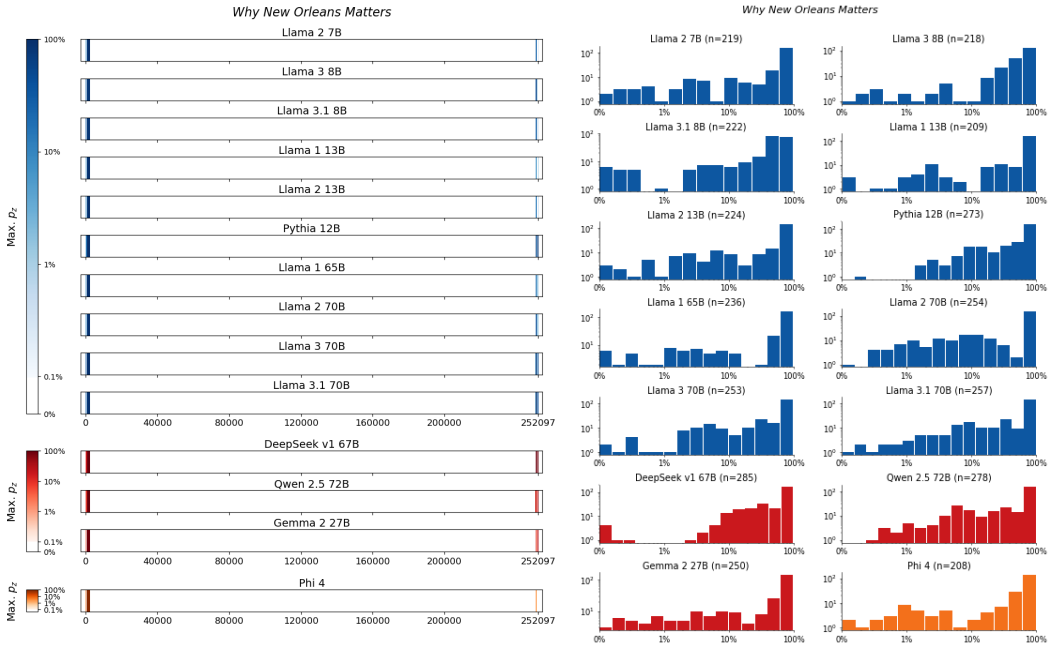


Figure 154: ***Why New Orleans Matters,* Piazza.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.144 *Enlightenment Now,* Pinker

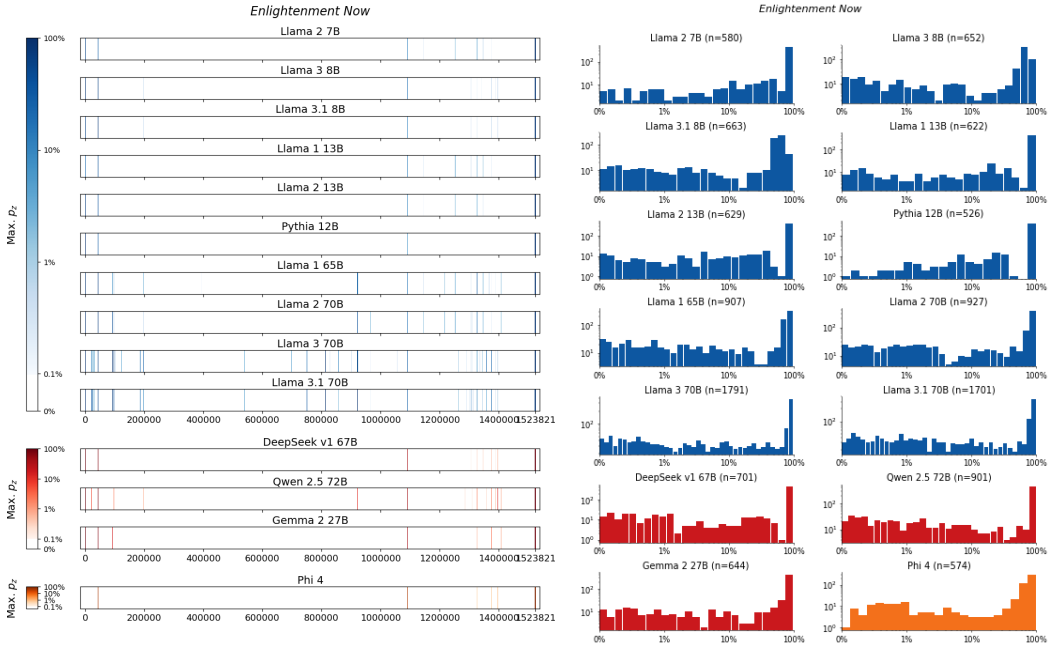


Figure 155: ***Enlightenment Now,* Pinker.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.145 *Competitive Strategy, Porter*

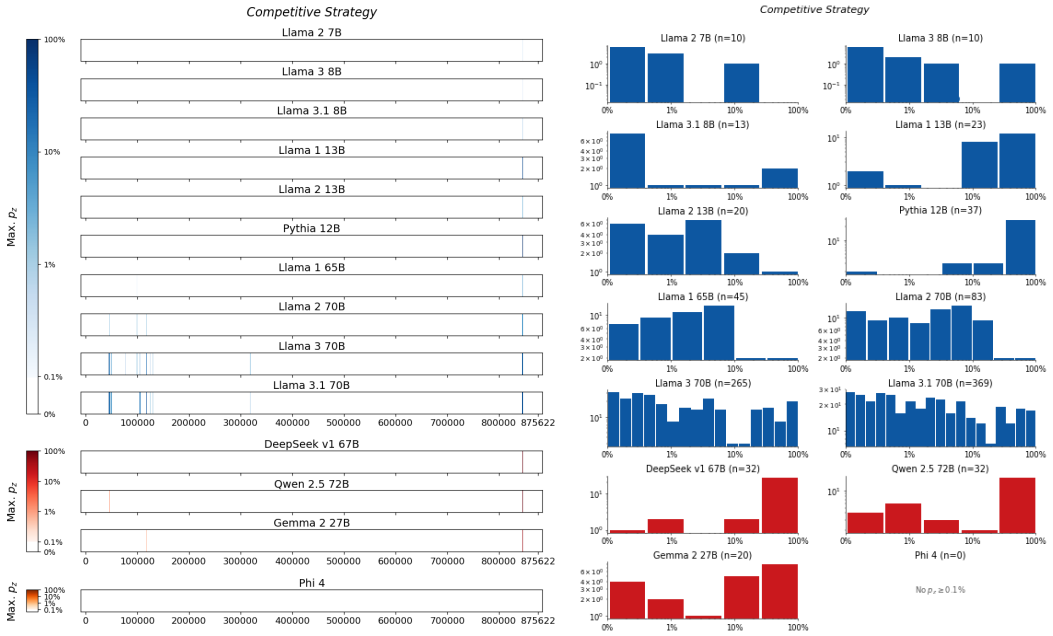


Figure 156: ***Competitive Strategy*, Porter.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.146 *Night Watch, Pratchett*

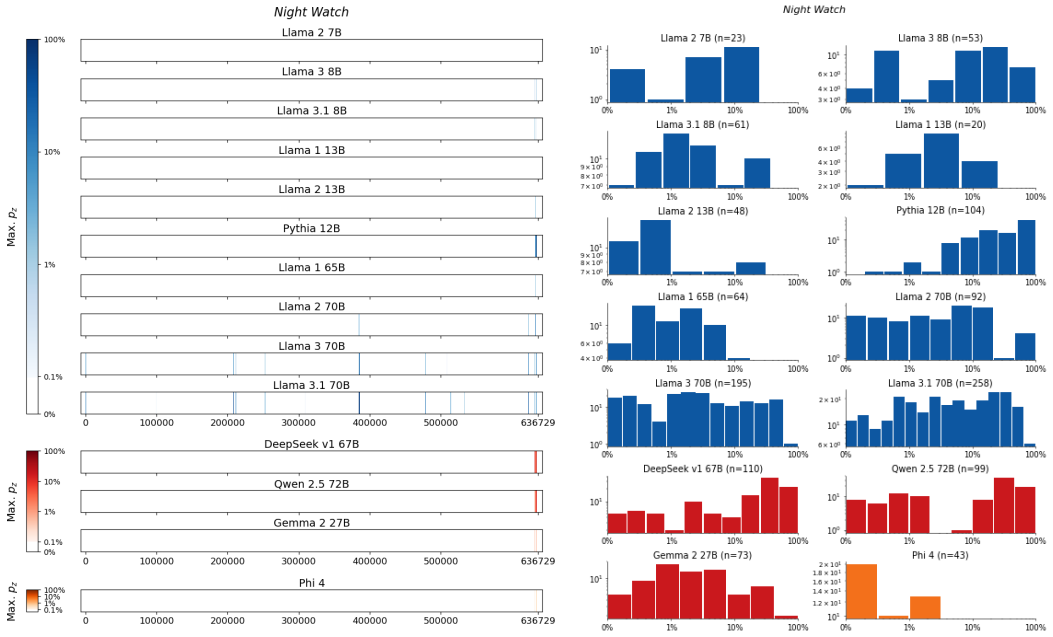


Figure 157: ***Night Watch*, Pratchett.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.147 The Subtle Knife, Pullman

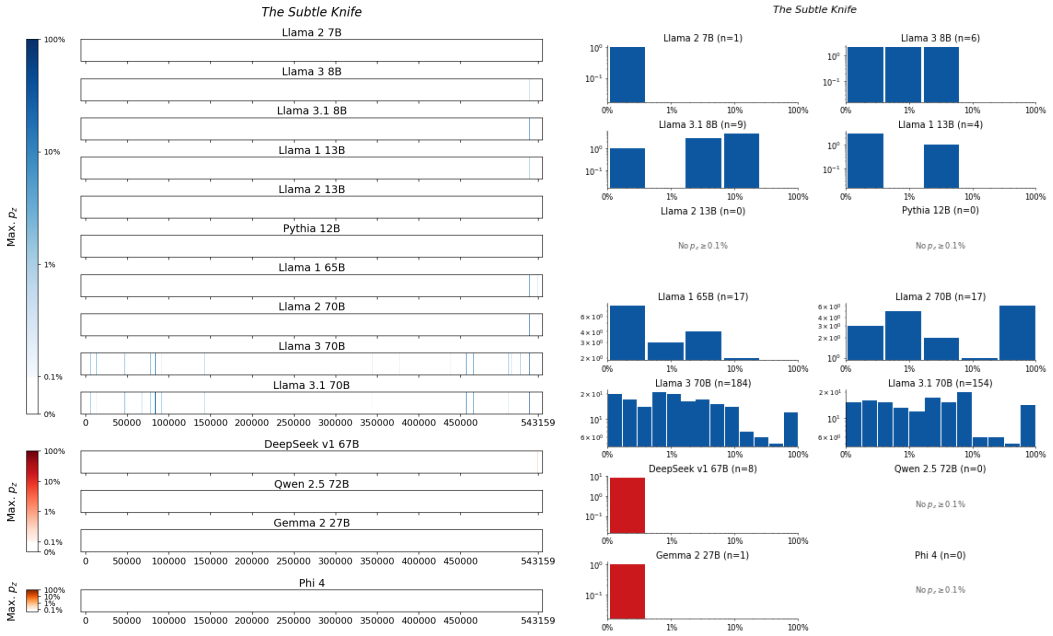


Figure 158: ***The Subtle Knife*, Pullman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.148 The Seductions of Darwin, Rampley

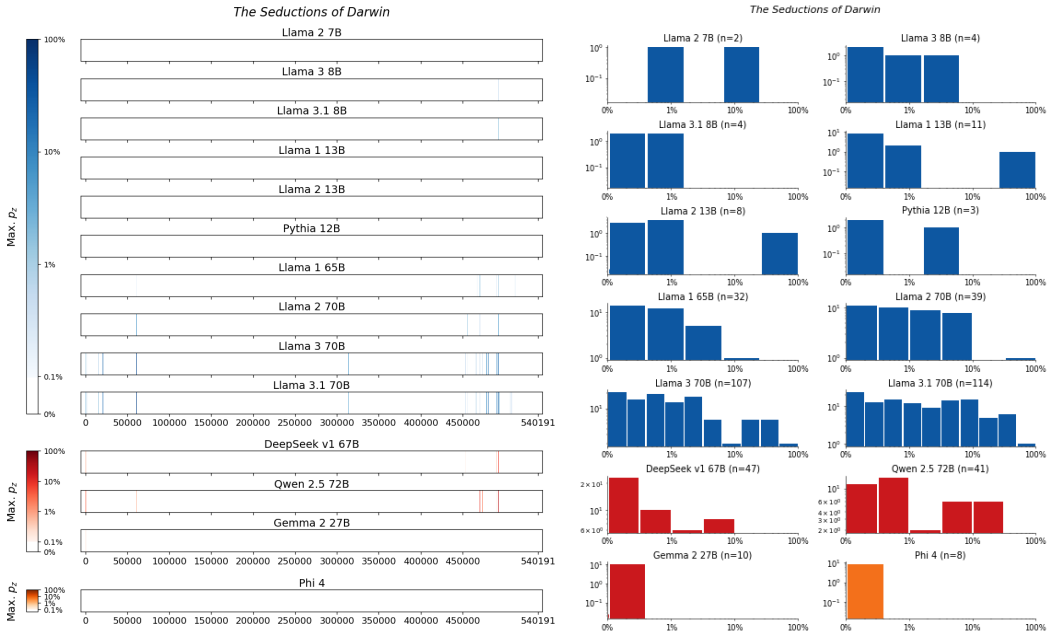


Figure 159: ***The Seductions of Darwin*, Rampley.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.149 *Kitchen Table Wisdom, Remen*

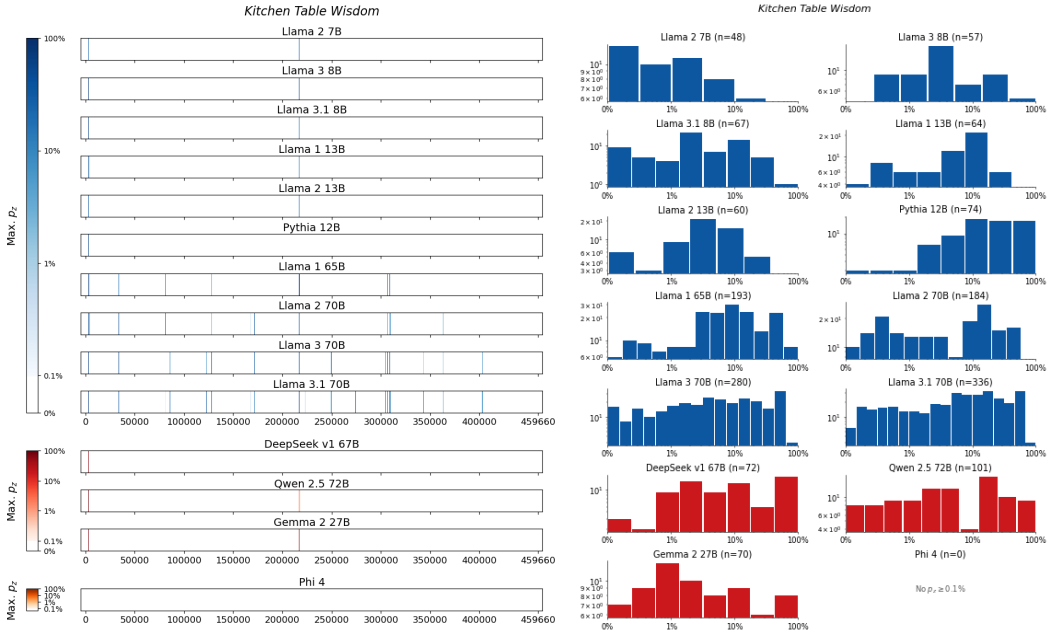


Figure 160: ***Kitchen Table Wisdom*, Remen.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.150 *Backroads Boss Lady, Roberts & Witter*

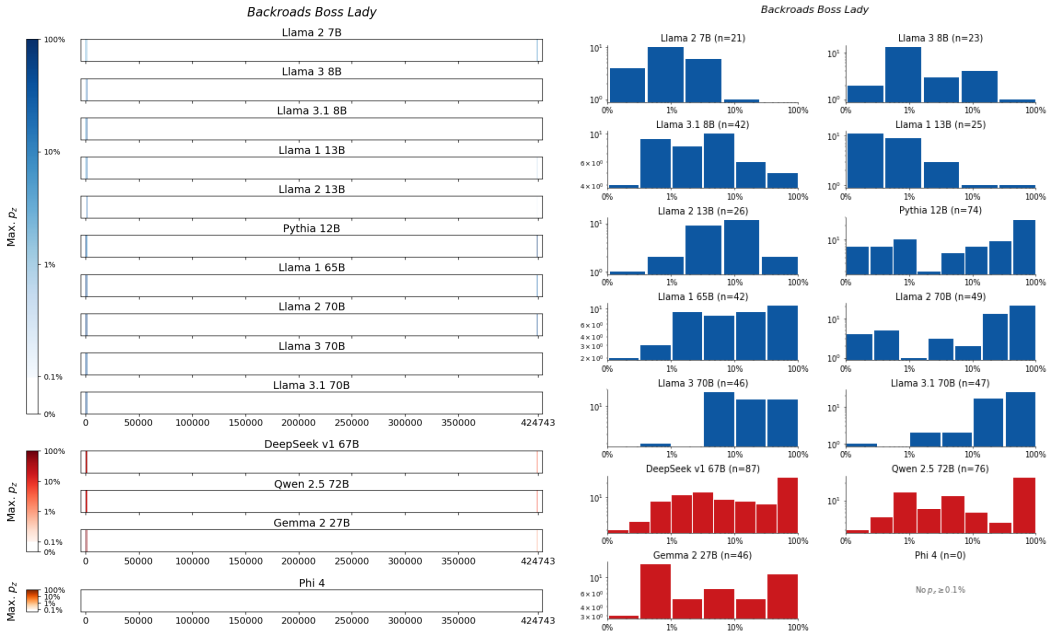


Figure 161: ***Backroads Boss Lady*, Roberts & Witter.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.151 *Soft in the Head,* Roger

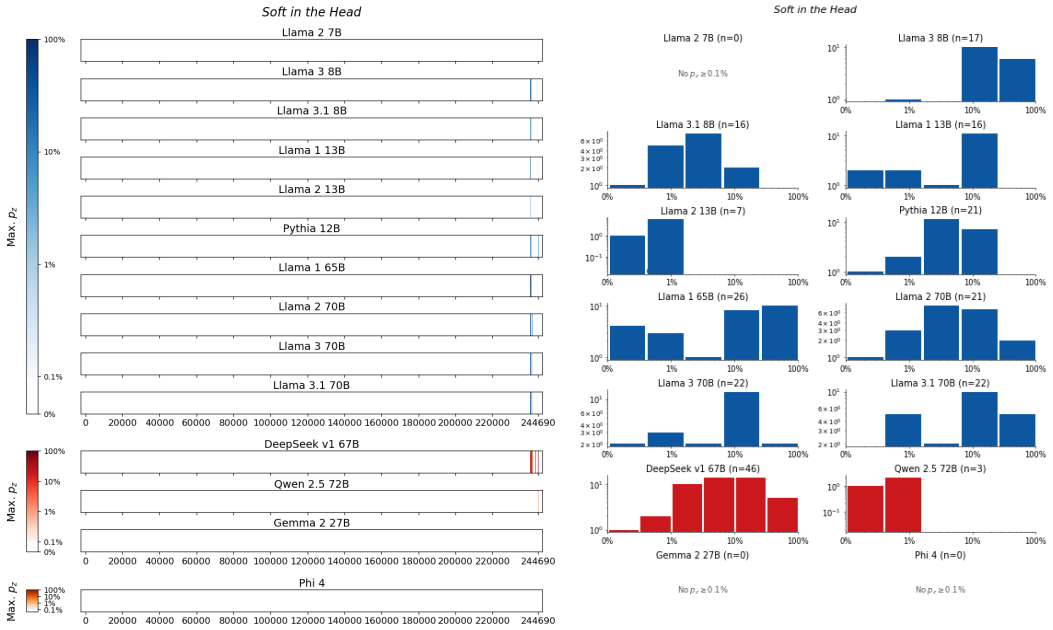


Figure 162: ***Soft in the Head,* Roger.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.152 *The Making of a Mediterranean Emirate,* Rouighi

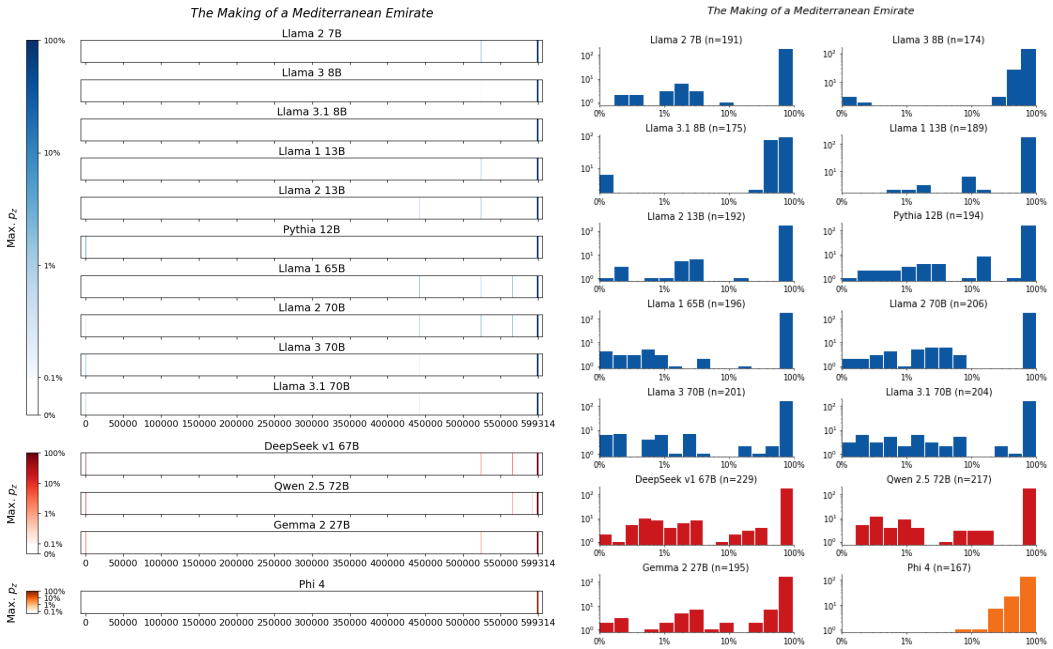


Figure 163: ***The Making of a Mediterranean Emirate,* Rouighi.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.153 *Harry Potter and the Sorcerer's Stone, Rowling*

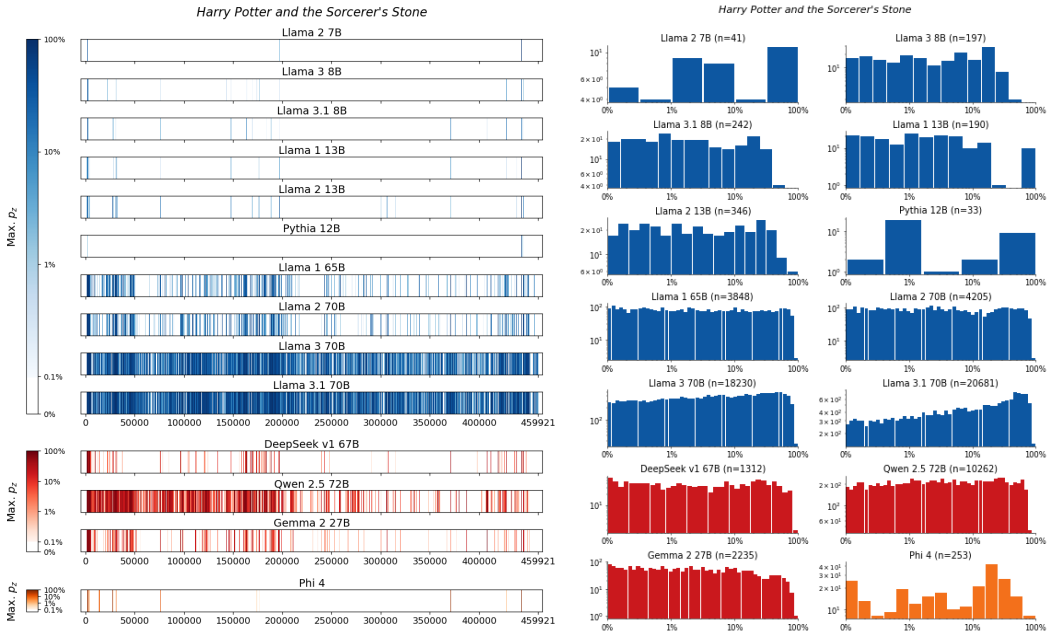


Figure 164: ***Harry Potter and the Sorcerer's Stone*, Rowling.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.154 *Harry Potter and the Chamber of Secrets, Rowling*

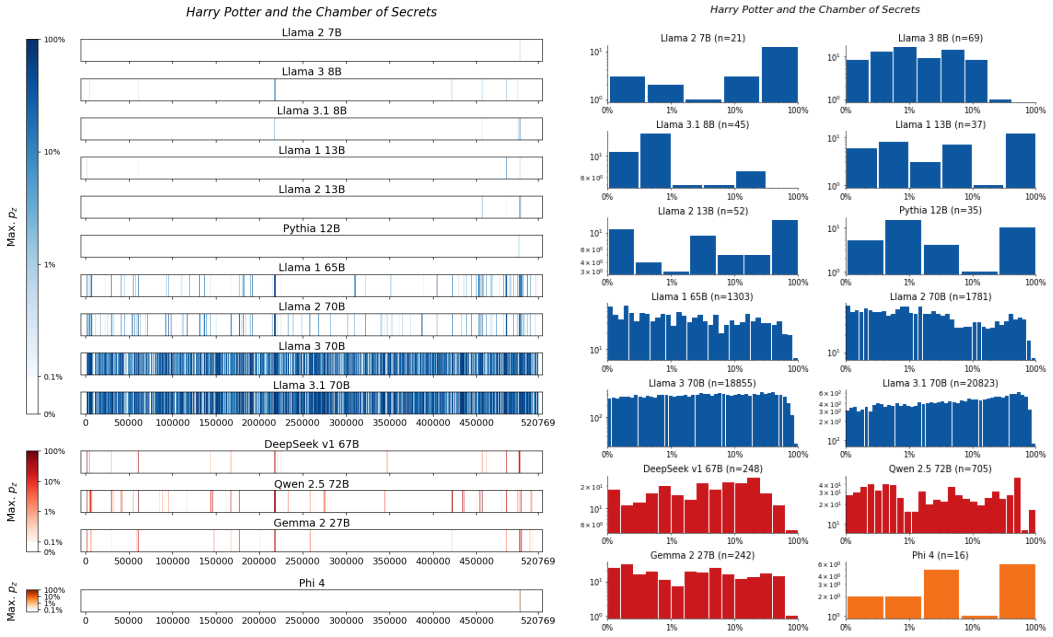


Figure 165: ***Harry Potter and the Chamber of Secrets*, Rowling.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.155 *Harry Potter and the Goblet of Fire,* Rowling

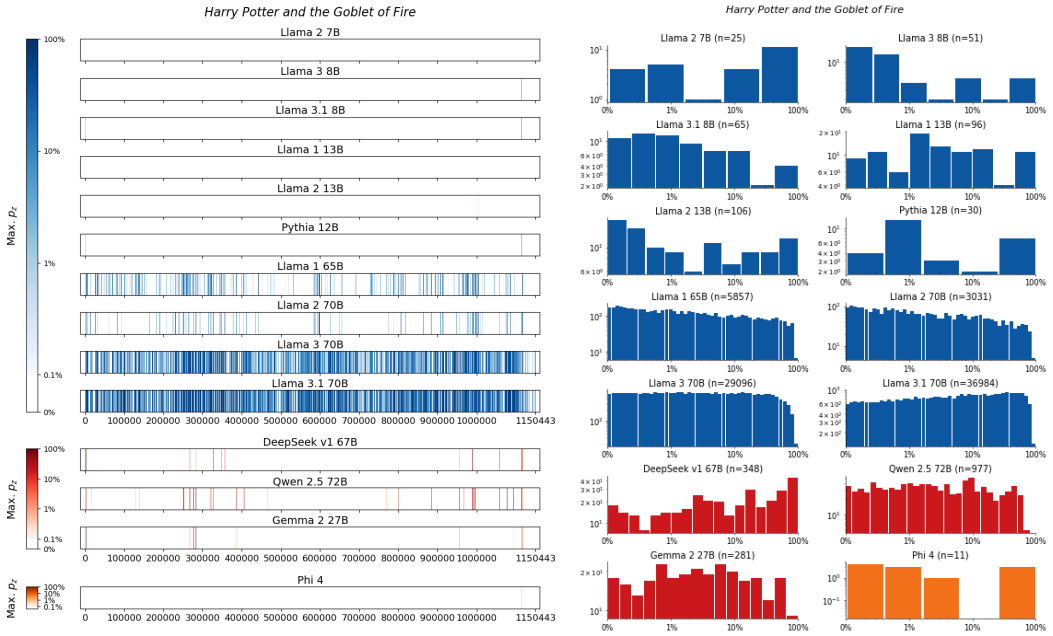


Figure 166: ***Harry Potter and the Goblet of Fire*, Rowling.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.156 *Harry Potter and the Deathly Hallows,* Rowling

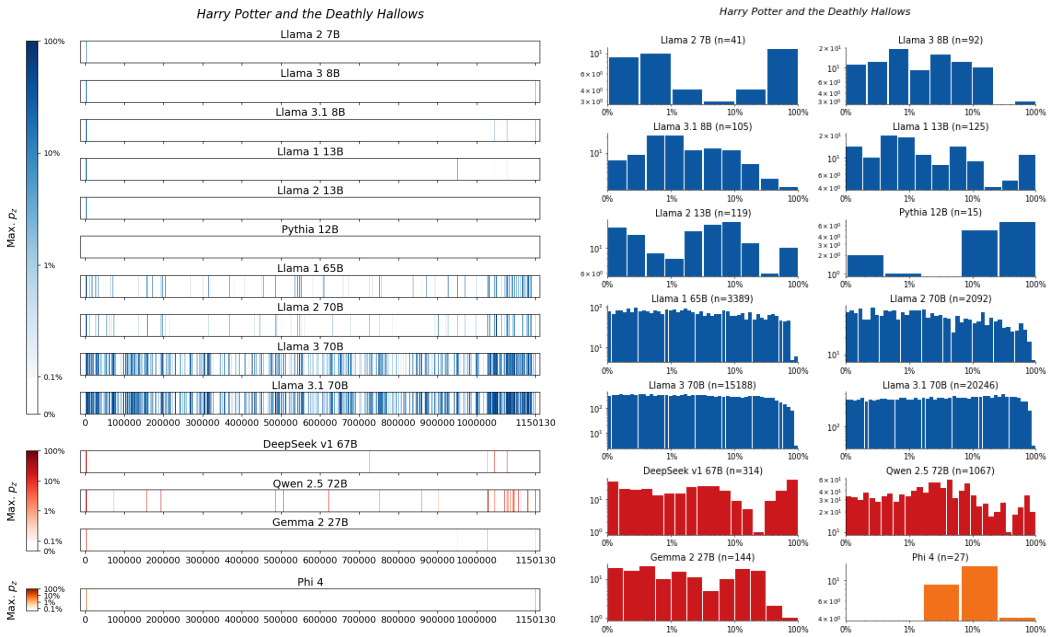


Figure 167: ***Harry Potter and the Deathly Hallows*, Rowling.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.157 Born to Walk, Rubinstein

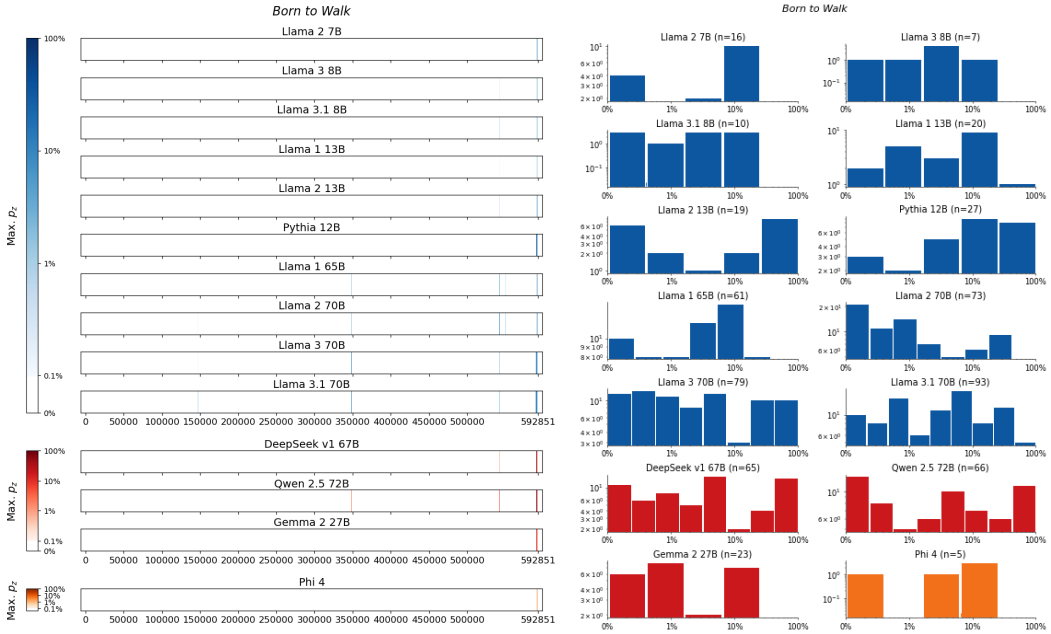


Figure 168: ***Born to Walk*, Rubinstein.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.158 The Pretender, Ruskin

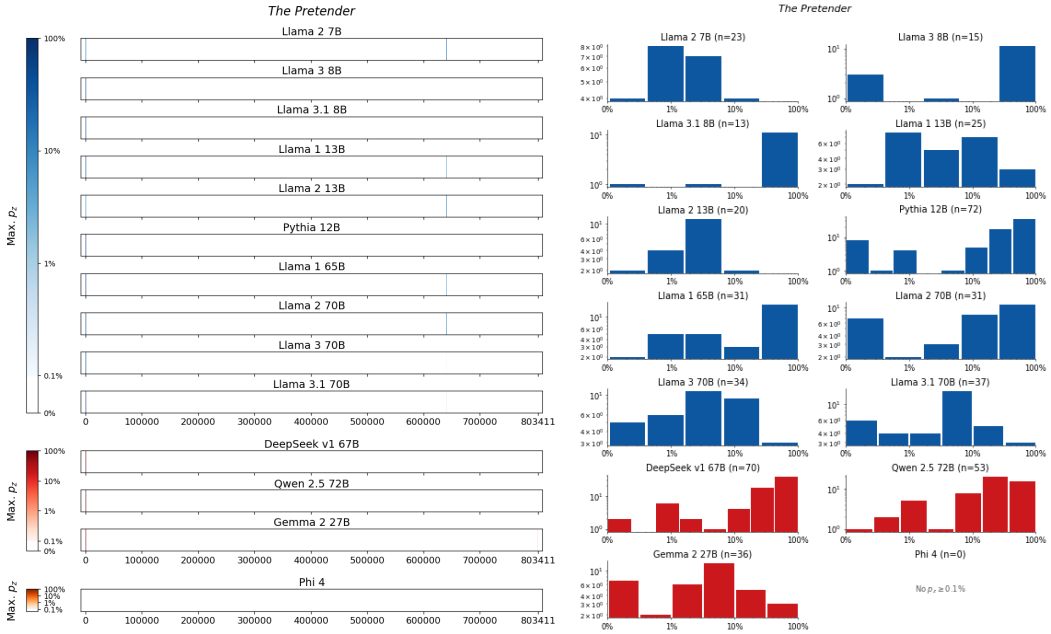


Figure 169: ***The Pretender*, Ruskin.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.159 Toscanini, Sachs

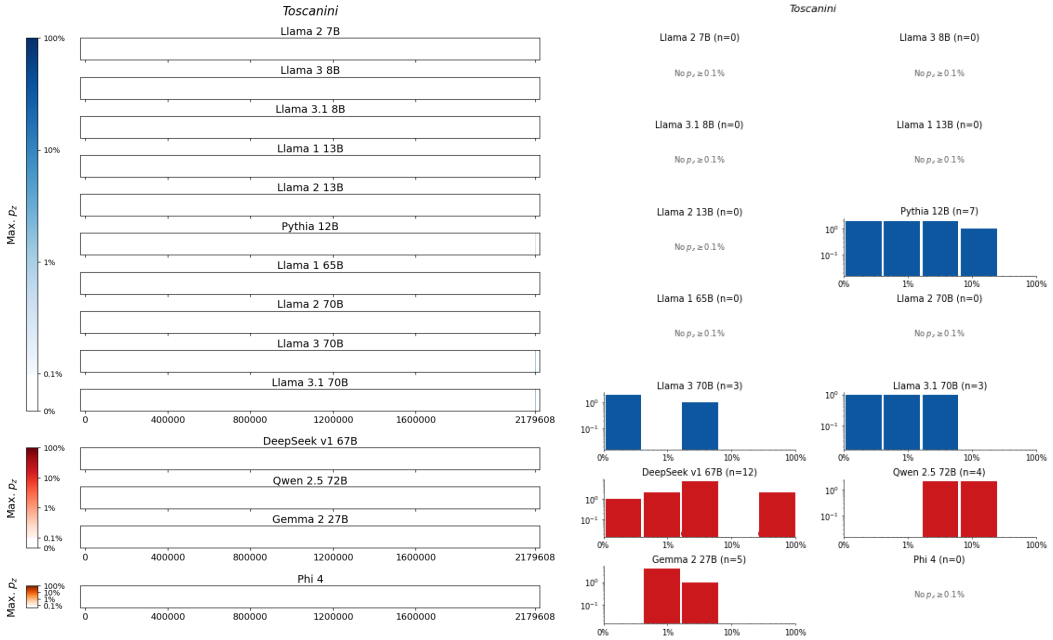


Figure 170: ***Toscanini*, Sachs.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.160 Cosmos, Sagan

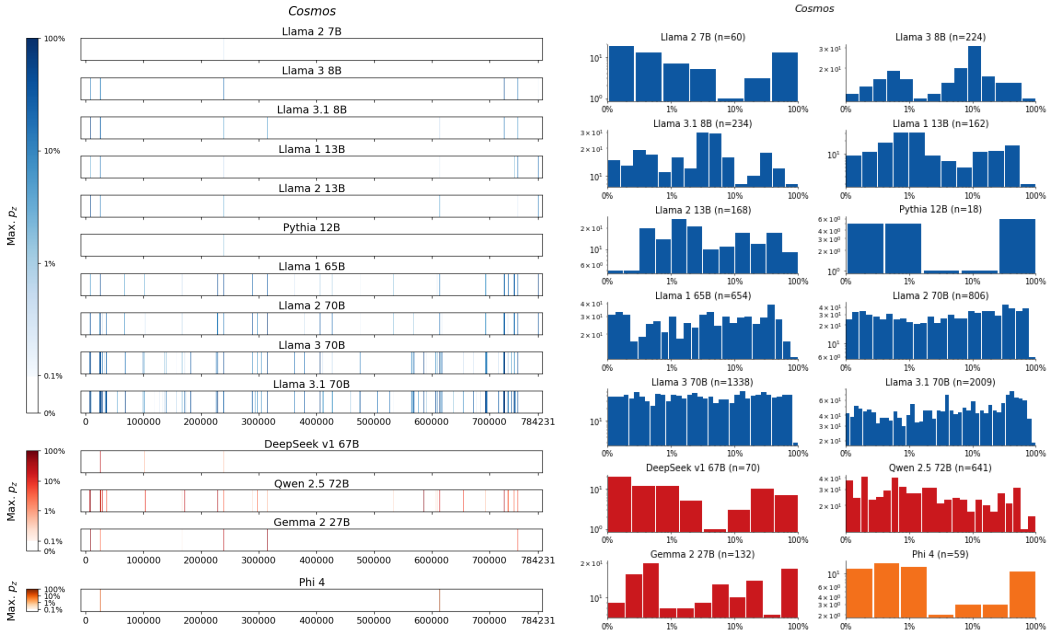


Figure 171: ***Cosmos*, Sagan.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.161 *Middle India,* Sahni

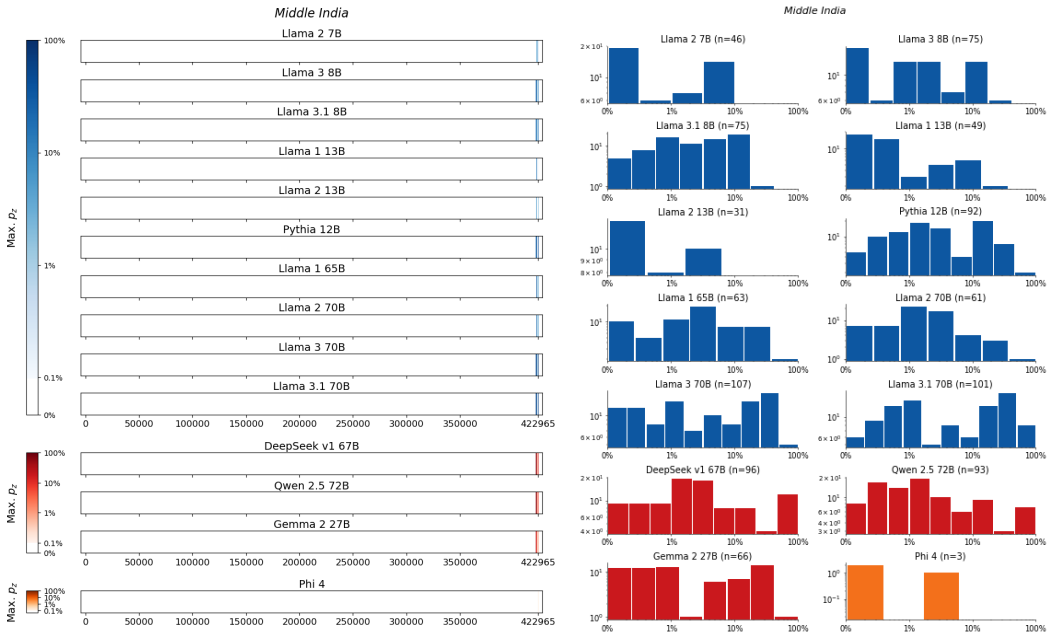


Figure 172: ***Middle India,* Sahni.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.162 *The Catcher in the Rye,* Salinger

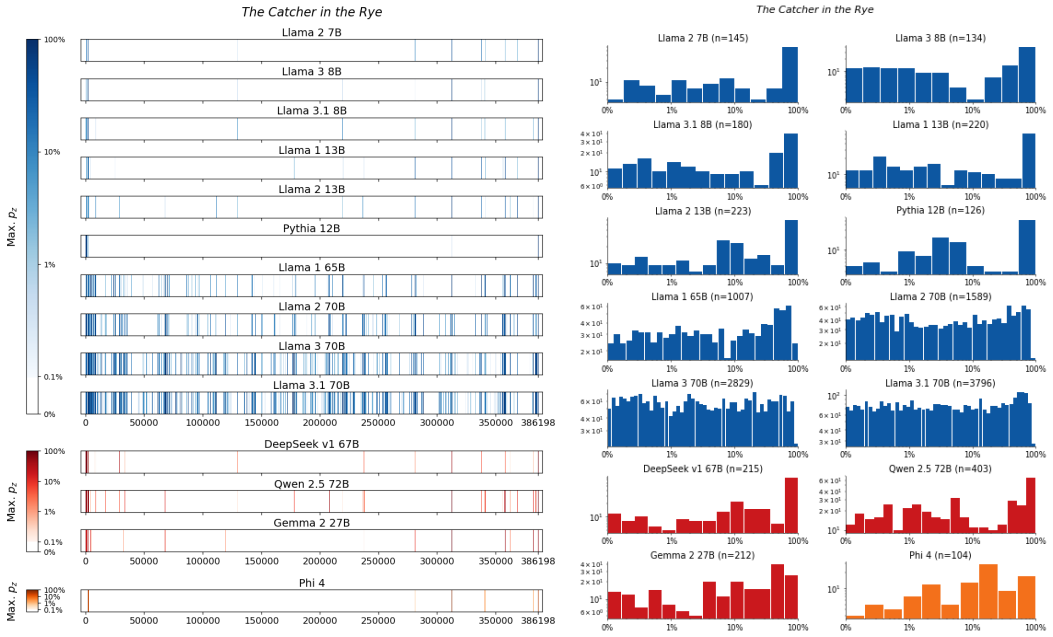


Figure 173: ***The Catcher in the Rye,* Salinger.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.163 Lean In, Sandberg

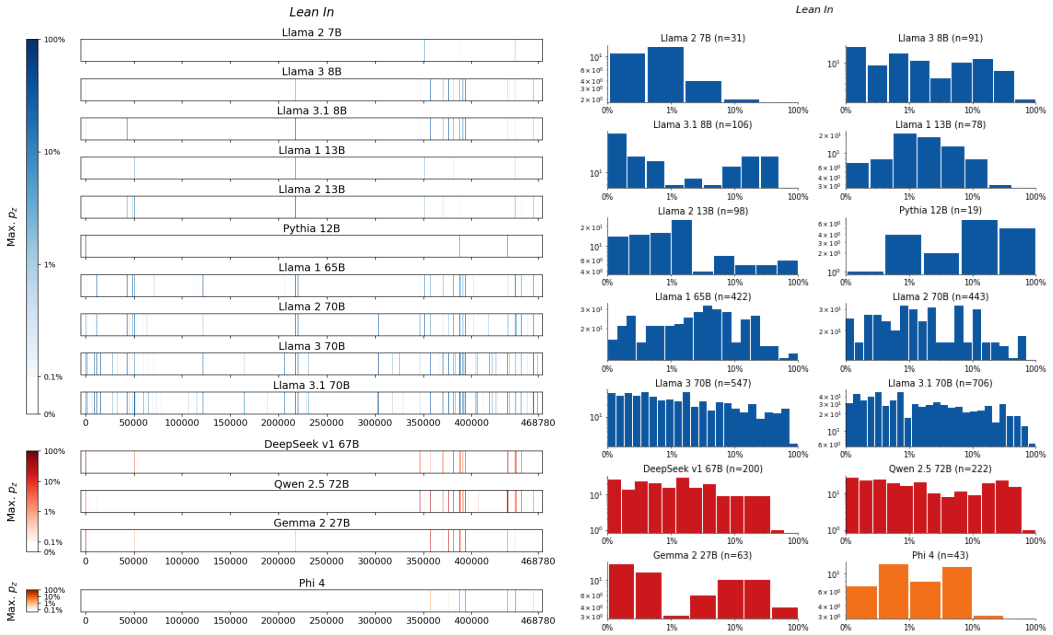


Figure 174: ***Lean In*, Sandberg.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.164 The DevOps Adoption Playbook, Sharma

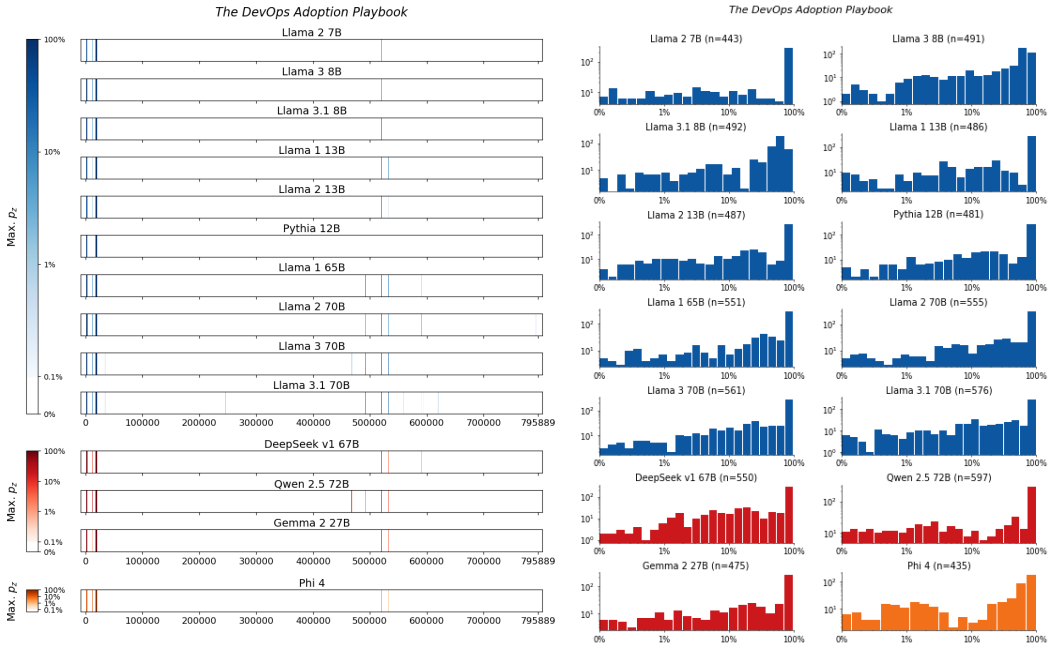


Figure 175: ***The DevOps Adoption Playbook*, Sharma.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.165 *Frankenstein*, *Shelley*

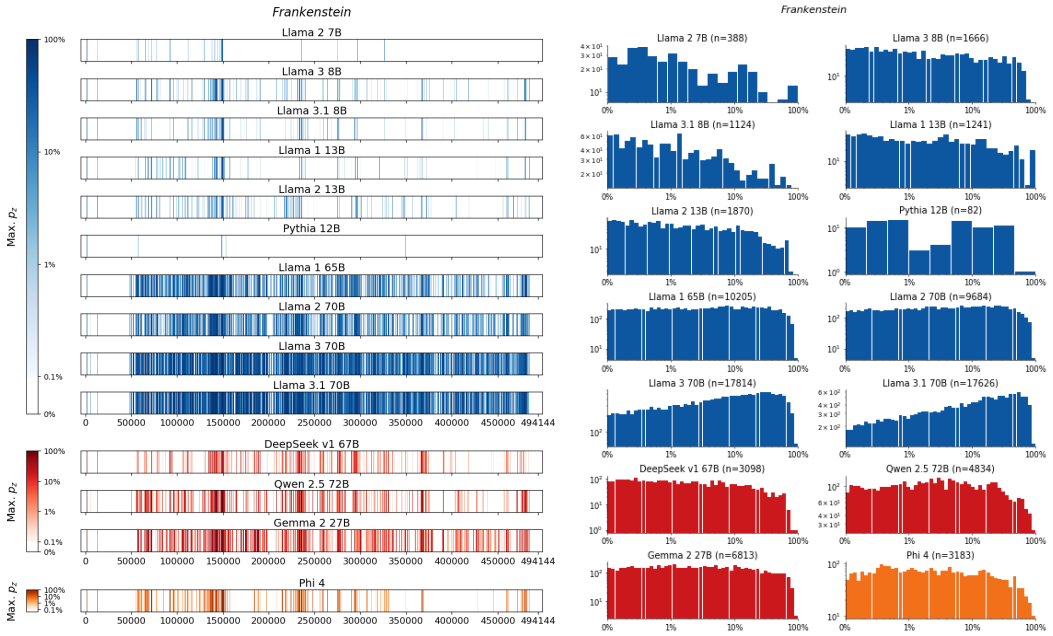


Figure 176: ***Frankenstein*, Shelley.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$). This is a public domain book, so PHI 4 is not a negative control.

E.3.166 *Sally Ride: America's First Woman in Space*, *Sherr*

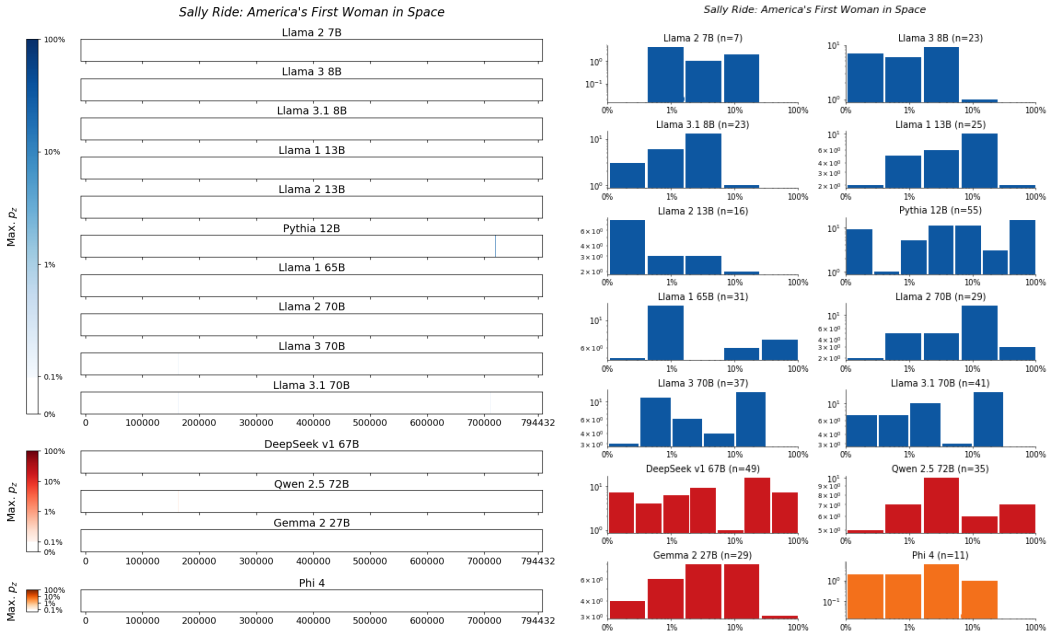


Figure 177: ***Sally Ride: America's First Woman in Space*, Sherr.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.167 A Perfectly Good Family, Shriver

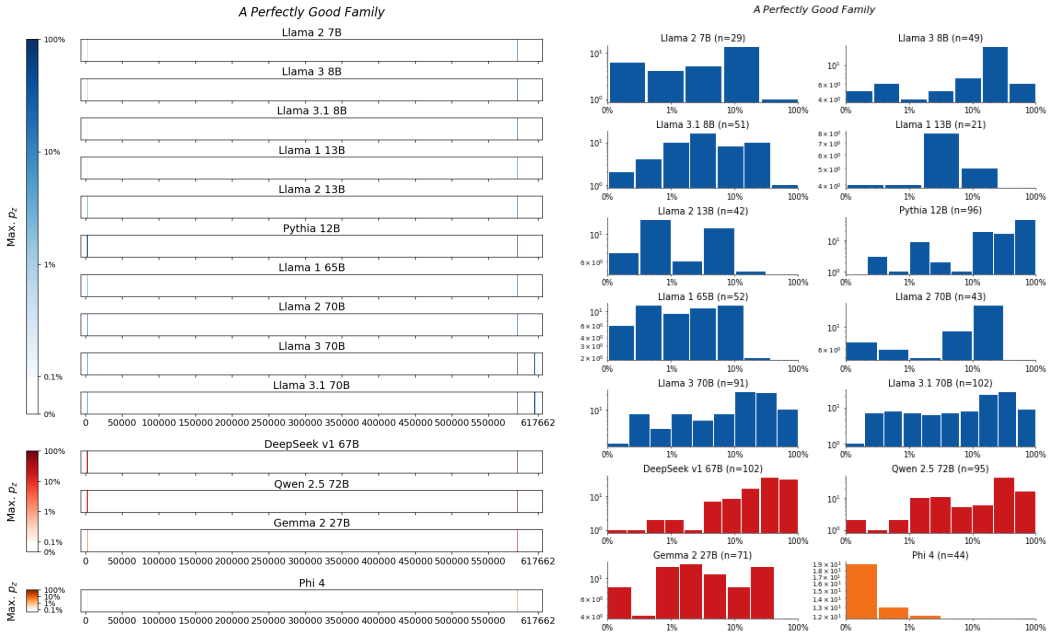


Figure 178: ***A Perfectly Good Family*, Shriver.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.168 The Bedwetter, Silverman

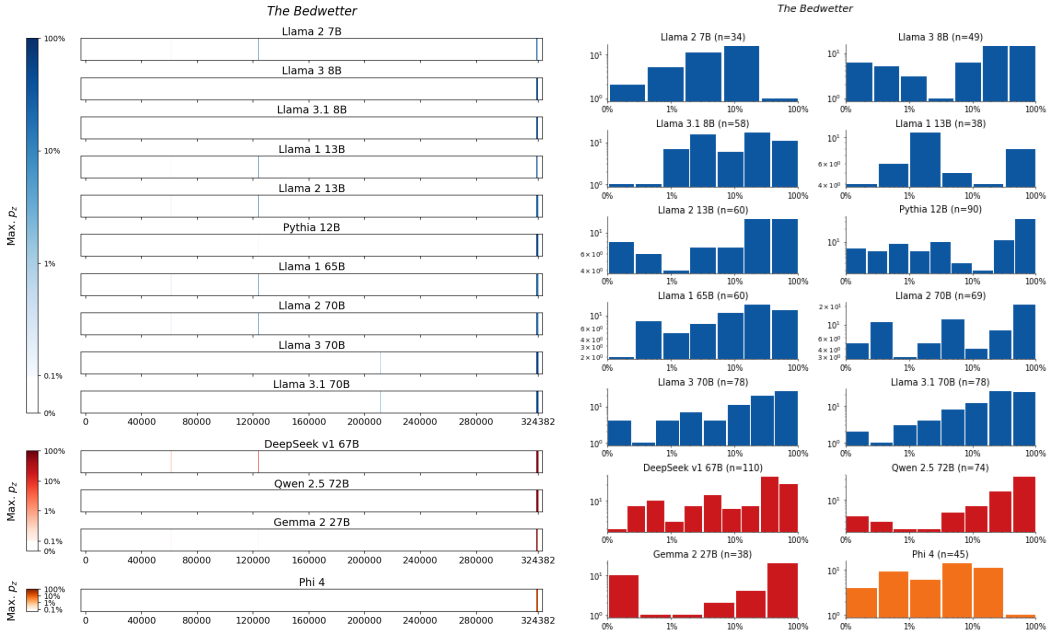


Figure 179: ***The Bedwetter*, Silverman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.169 On the Road with Bob Dylan, Sloman

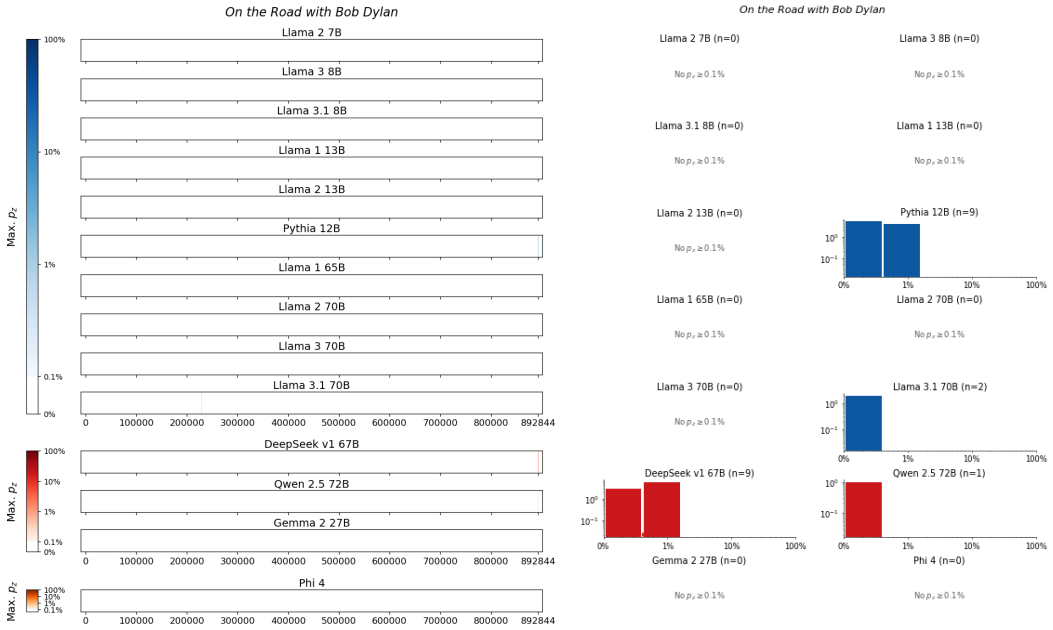


Figure 180: ***On the Road with Bob Dylan*, Sloman.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.170 The Night Children, Smith

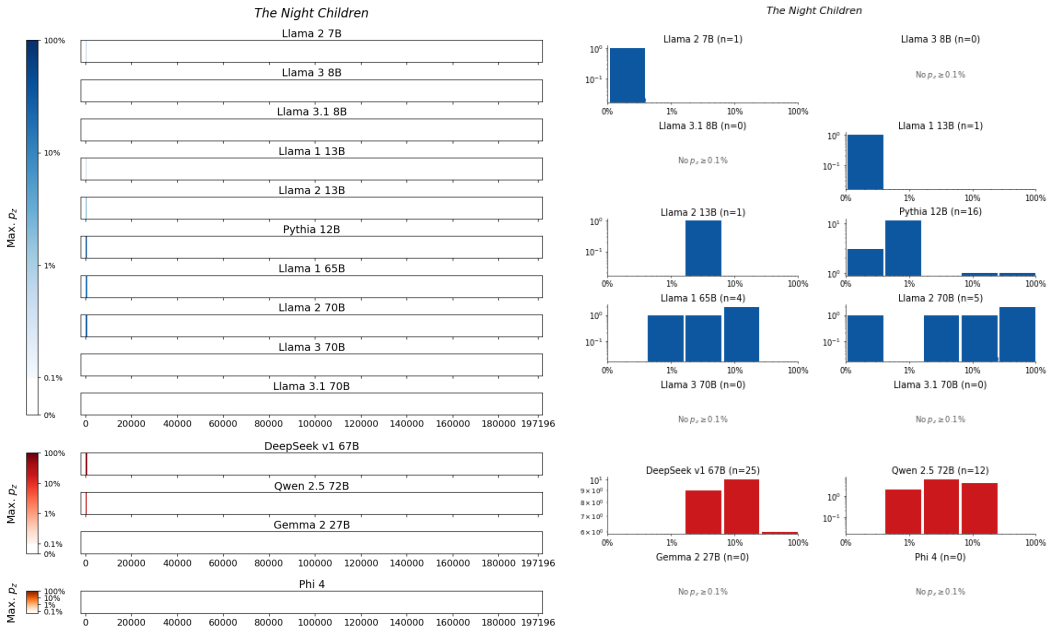


Figure 181: ***The Night Children*, Smith.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.171 White Teeth, Smith

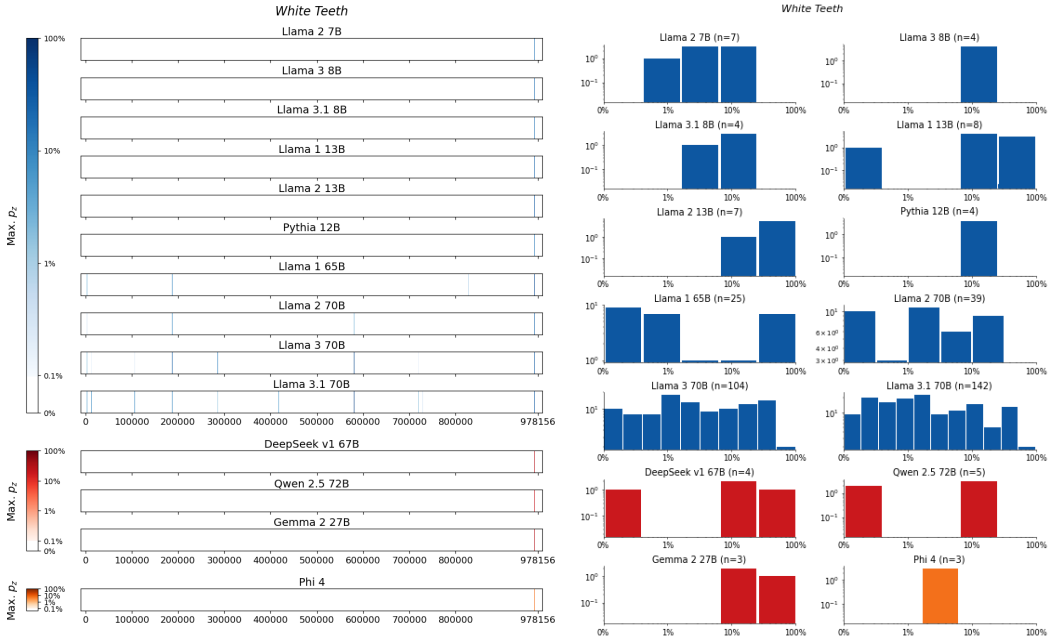


Figure 182: ***White Teeth*, Smith.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.172 No Visible Bruises, Snyder

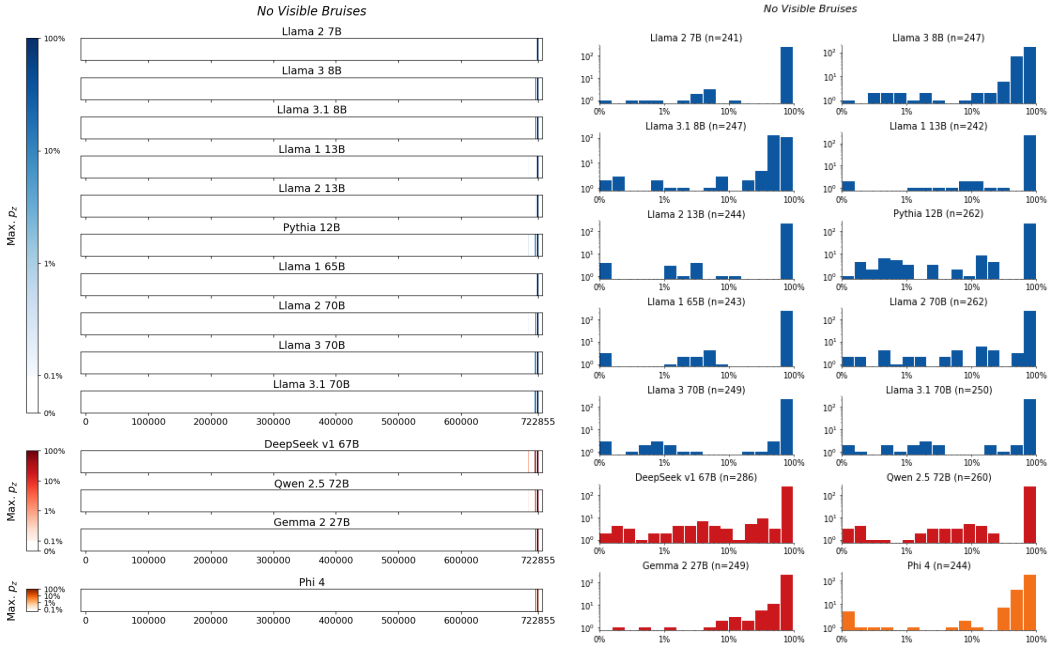


Figure 183: ***No Visible Bruises*, Snyder.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.173 *The Grapes of Wrath*, *Steinbeck*

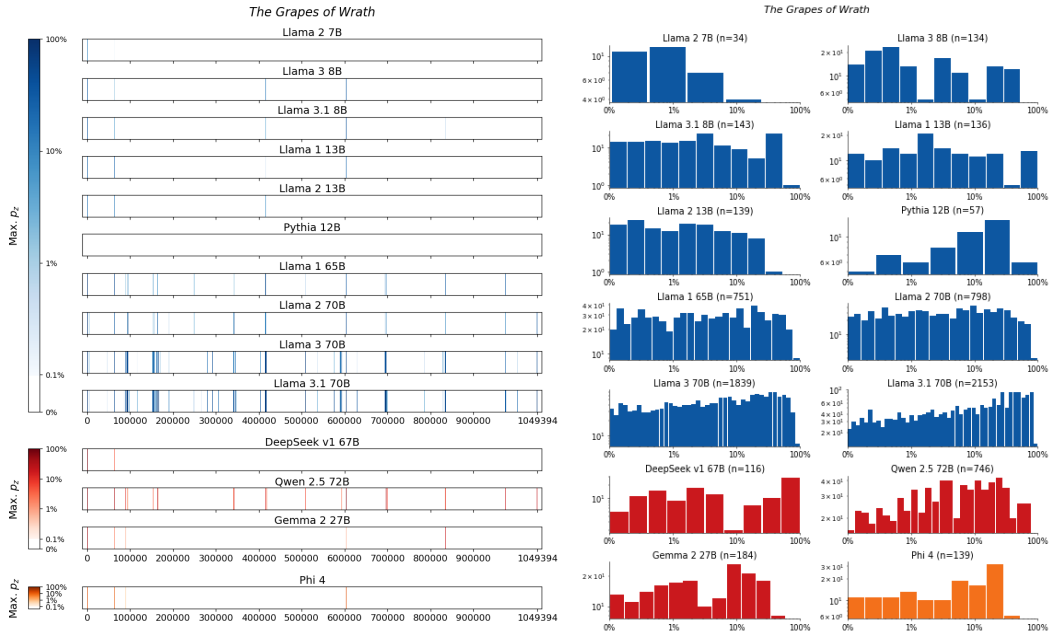


Figure 184: ***The Grapes of Wrath*, Steinbeck.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.174 *Jesse James*, *Stiles*

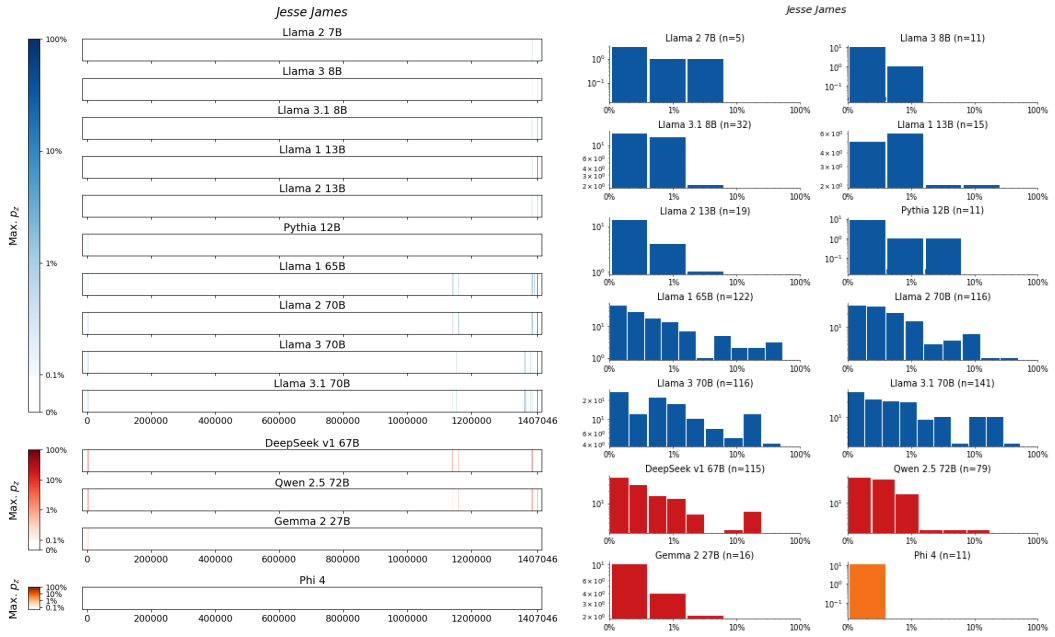


Figure 185: ***Jesse James*, Stiles.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.175 *Zombie Halloween,* Stine

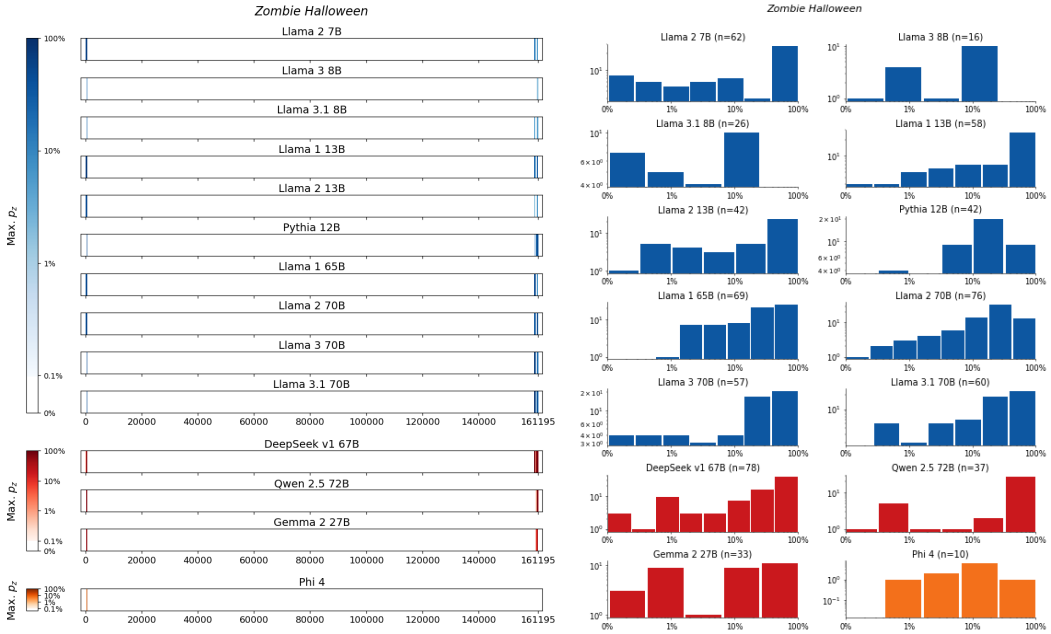


Figure 186: ***Zombie Halloween,* Stine.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.176 *Fear of Music,* Stubbs

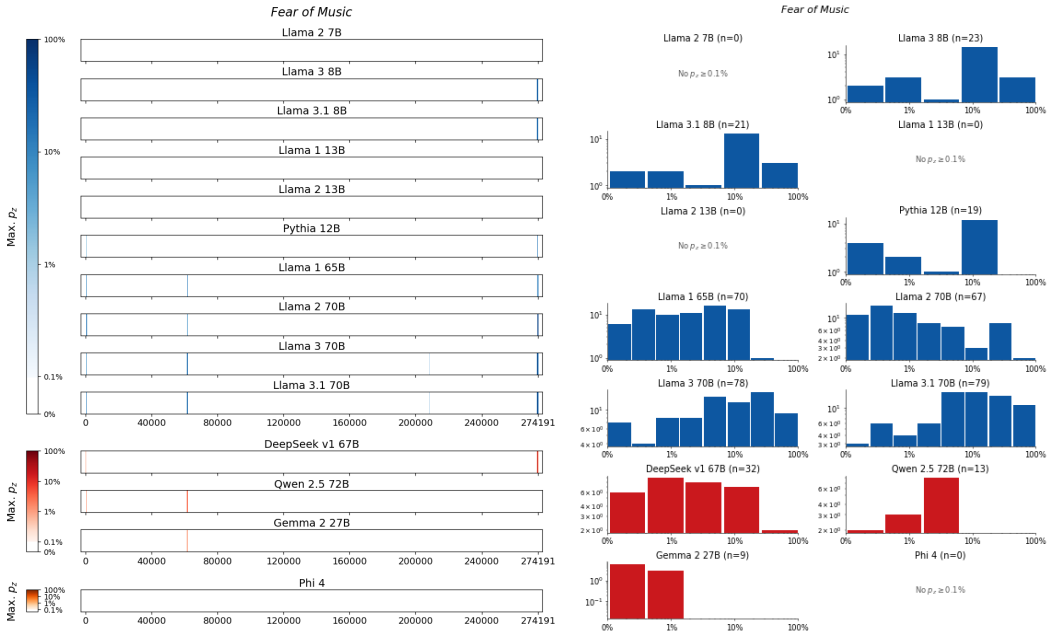


Figure 187: ***Fear of Music,* Stubbs.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.177 Pearl Harbor, Swanson

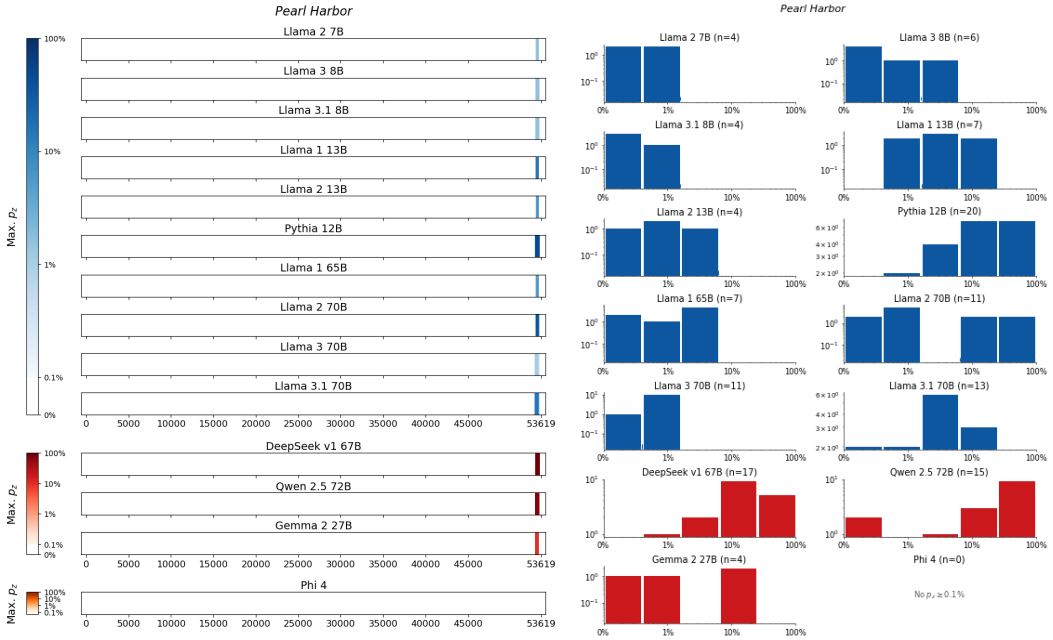


Figure 188: ***Pearl Harbor*, Swanson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.178 Chesapeake Requiem, Swift

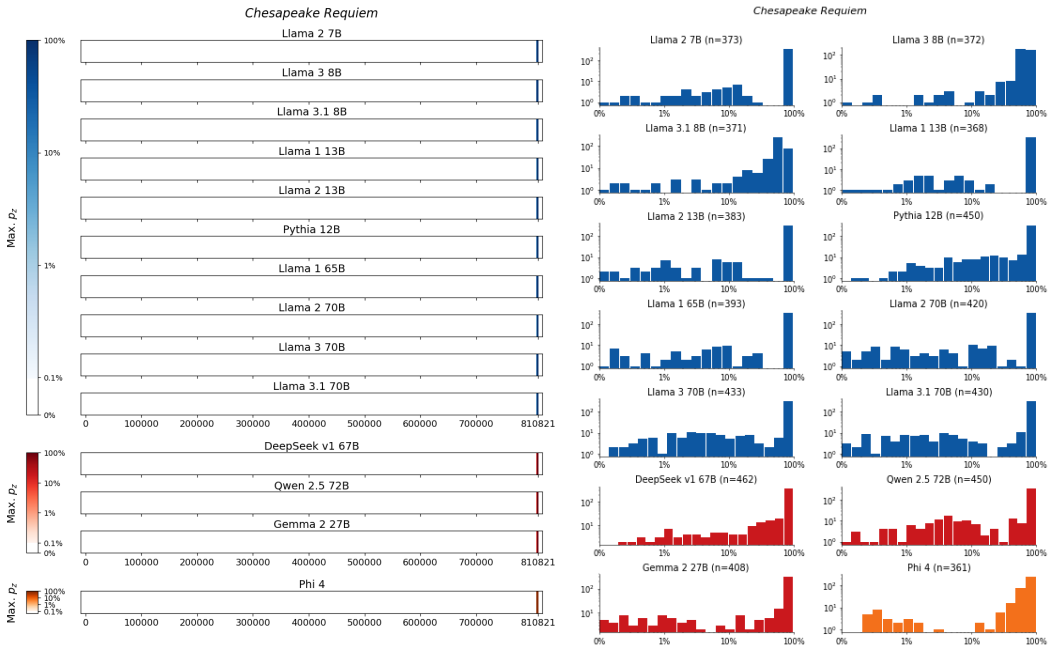


Figure 189: ***Chesapeake Requiem*, Swift.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.179 *The Goldfinch, Tartt*

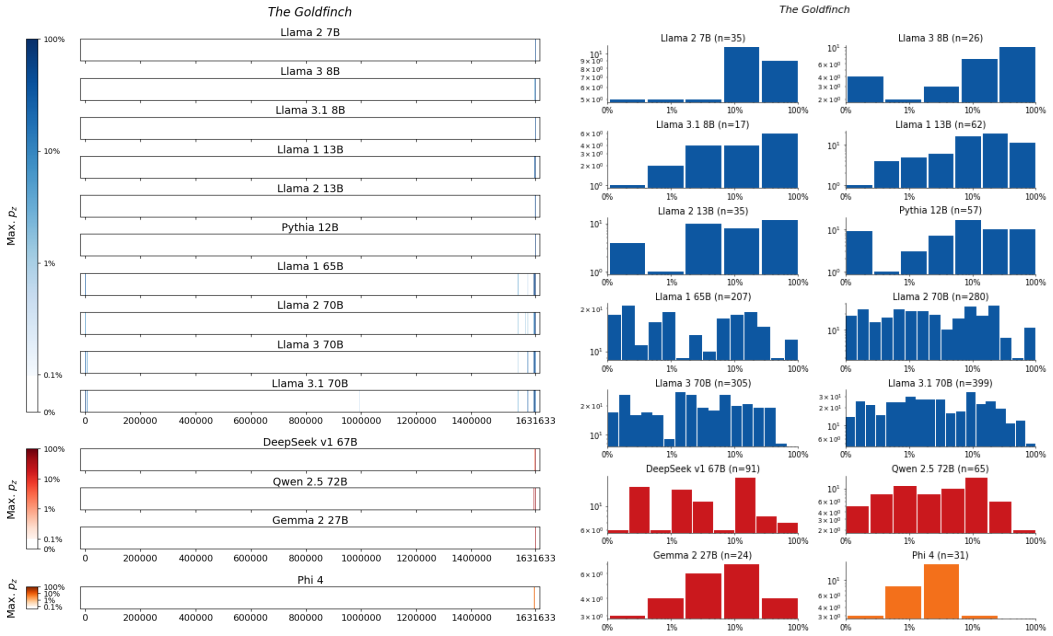


Figure 190: ***The Goldfinch*, Tartt.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.180 *Unglued, TerKeurst*

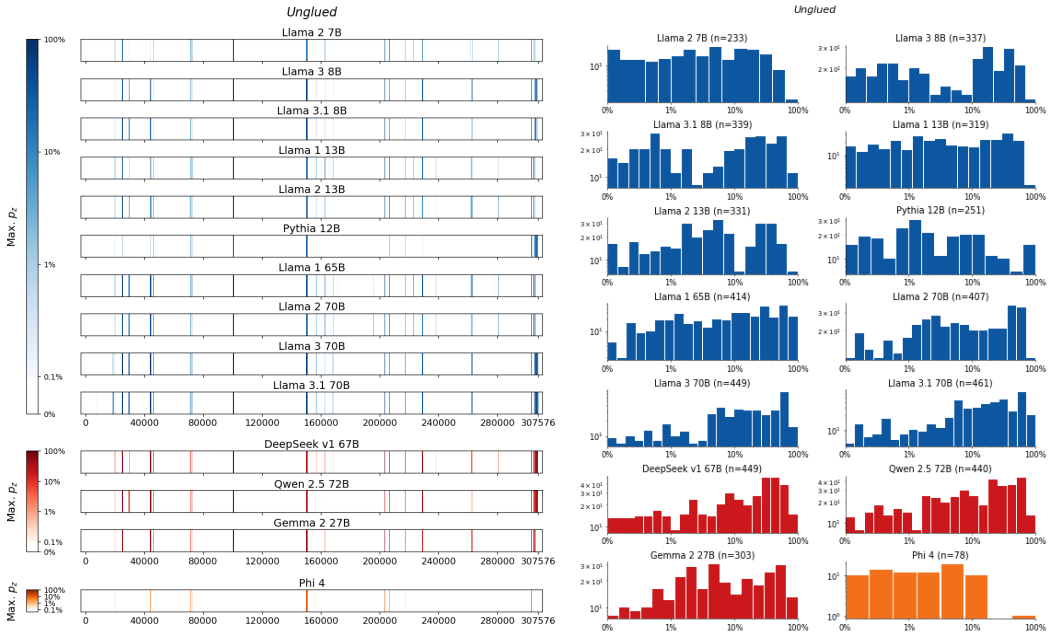


Figure 191: ***Unglued*, TerKeurst.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.181 *Embraced, TerKeurst*

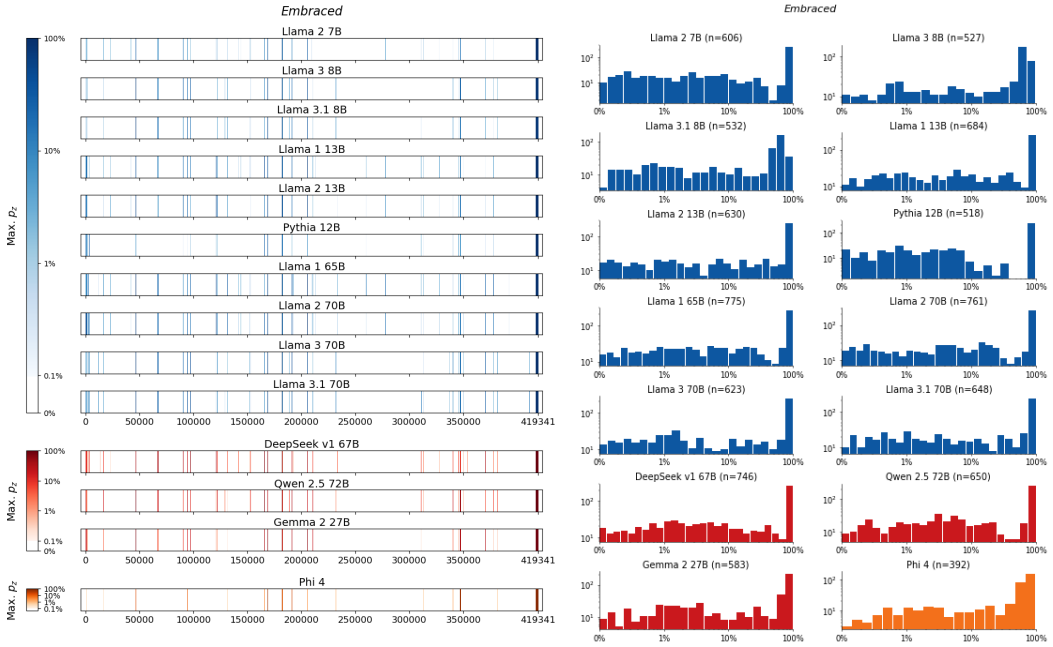


Figure 192: ***Embraced, TerKeurst.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.182 *Birding with Yeats, Thomson*

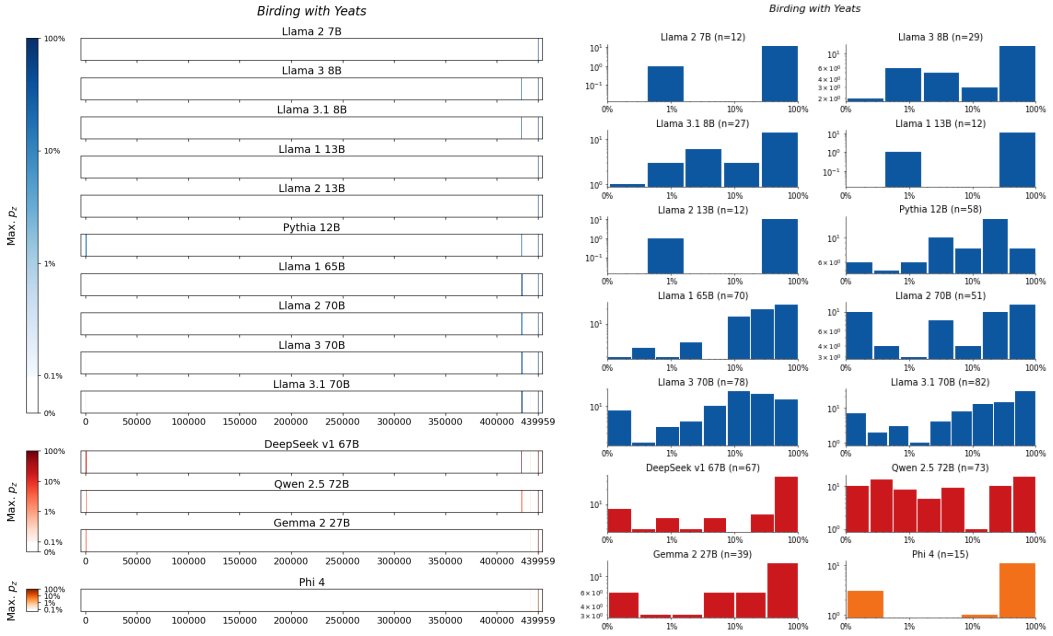


Figure 193: ***Birding with Yeats, Thomson.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.183 The Hobbit, Tolkien

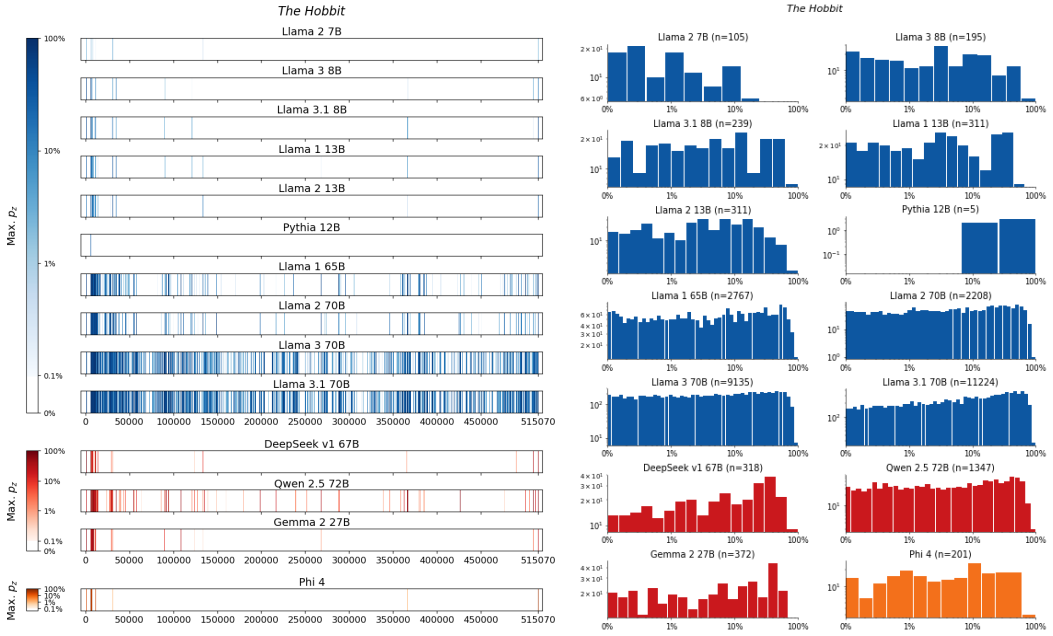


Figure 194: ***The Hobbit*, Tolkien.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.184 The Fellowship of the Ring, Tolkien

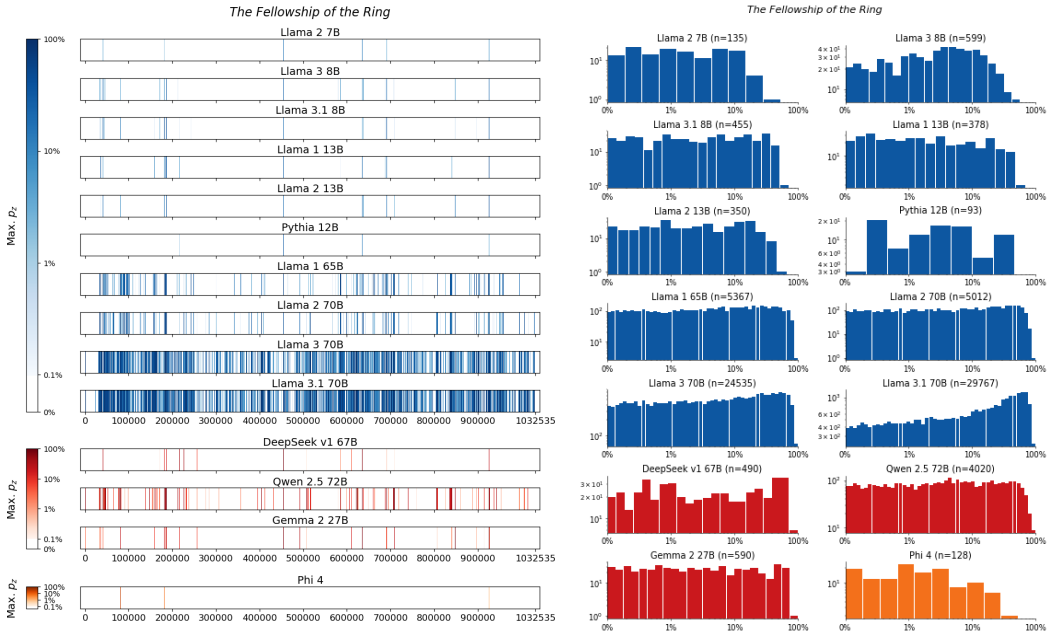


Figure 195: ***The Fellowship of the Ring*, Tolkien.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.185 Tree and Leaf, Tolkien

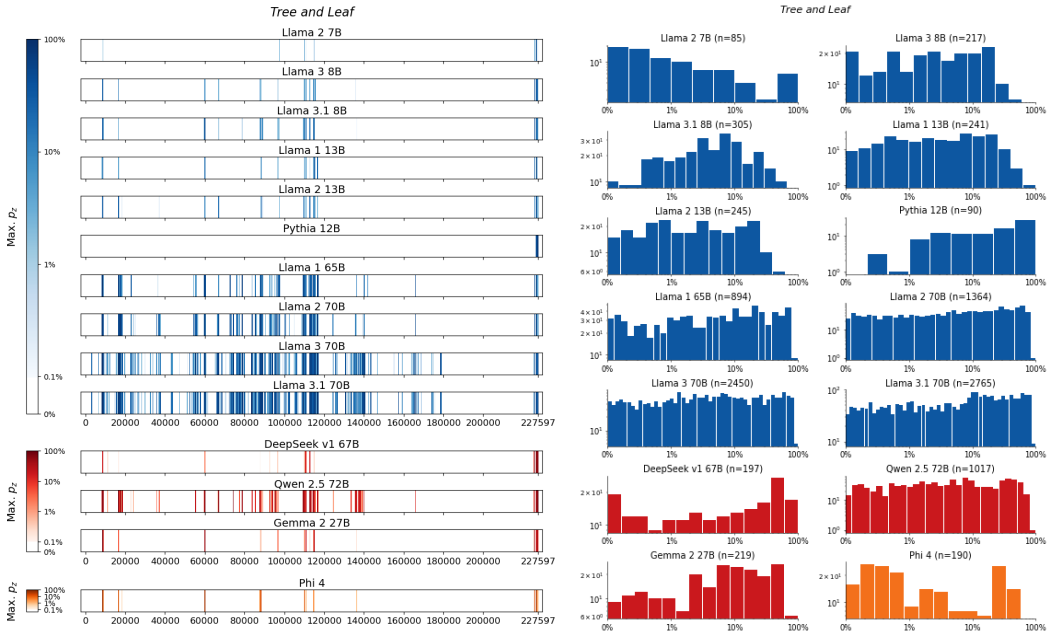


Figure 196: ***Tree and Leaf*, Tolkien.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.186 Noodles Every Day, Trang

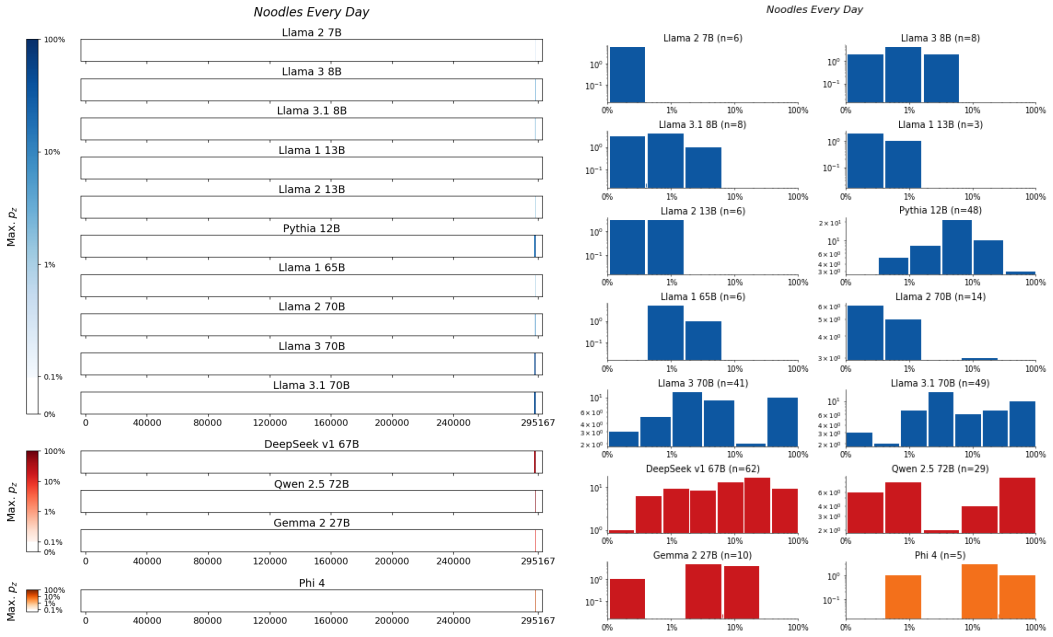


Figure 197: ***Noodles Every Day*, Trang.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.187 *Billionaire Democracy, Tyler*

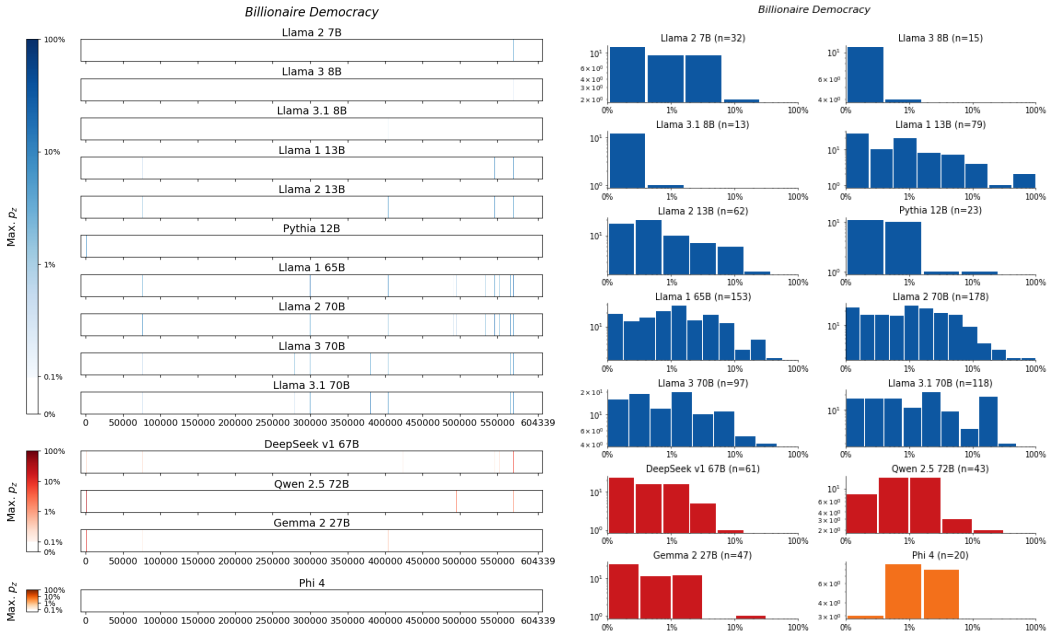


Figure 198: ***Billionaire Democracy*, Tyler.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.188 *Portugal's Guerrilla Wars in Africa, Venter*

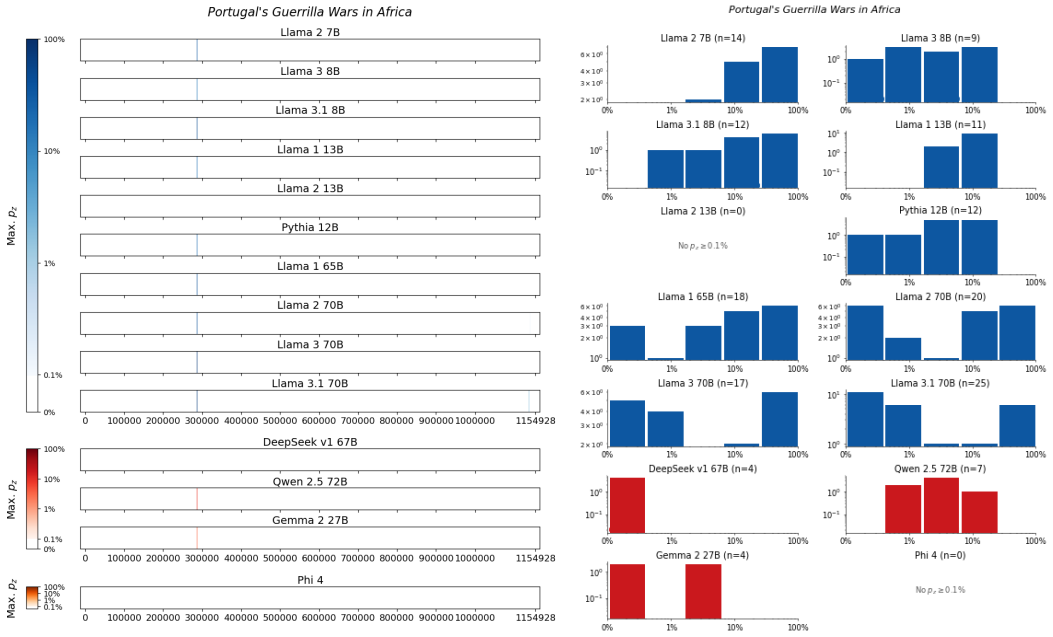


Figure 199: ***Portugal's Guerrilla Wars in Africa*, Venter.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.189 Slaughterhouse-Five, Vonnegut

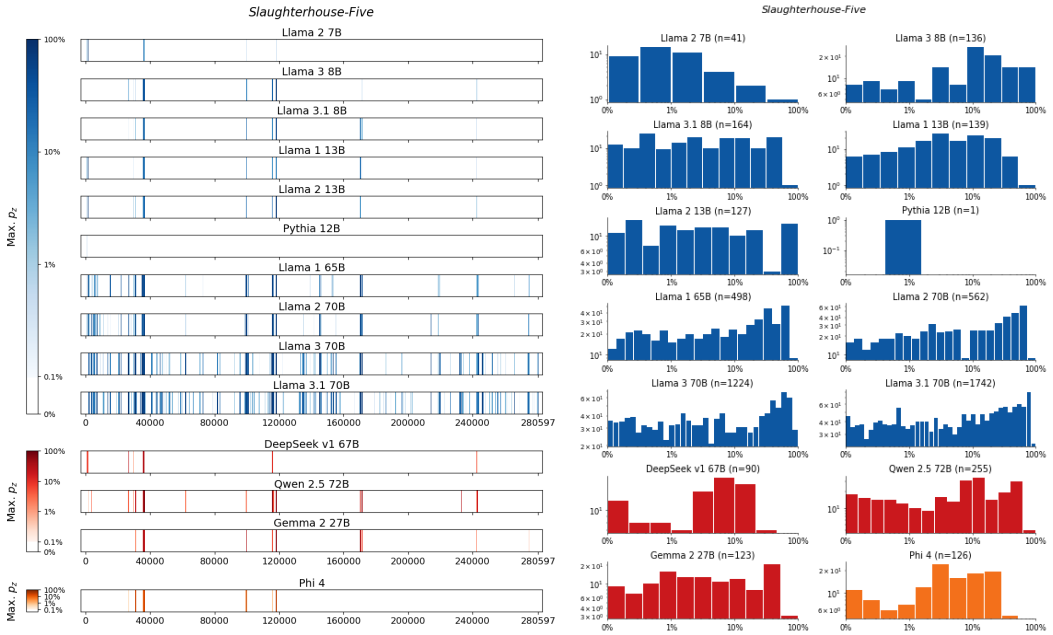


Figure 200: ***Slaughterhouse-Five*, Vonnegut.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.190 Animal Rights, Waldau

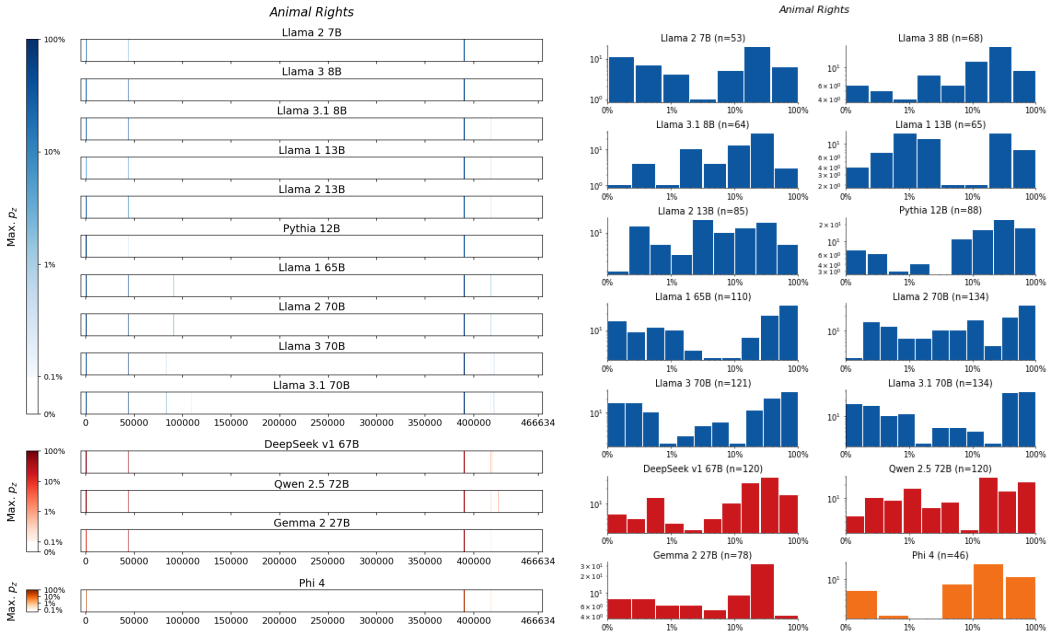


Figure 201: ***Animal Rights*, Waldau.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.191 *Men We Reaped,* ***Ward***

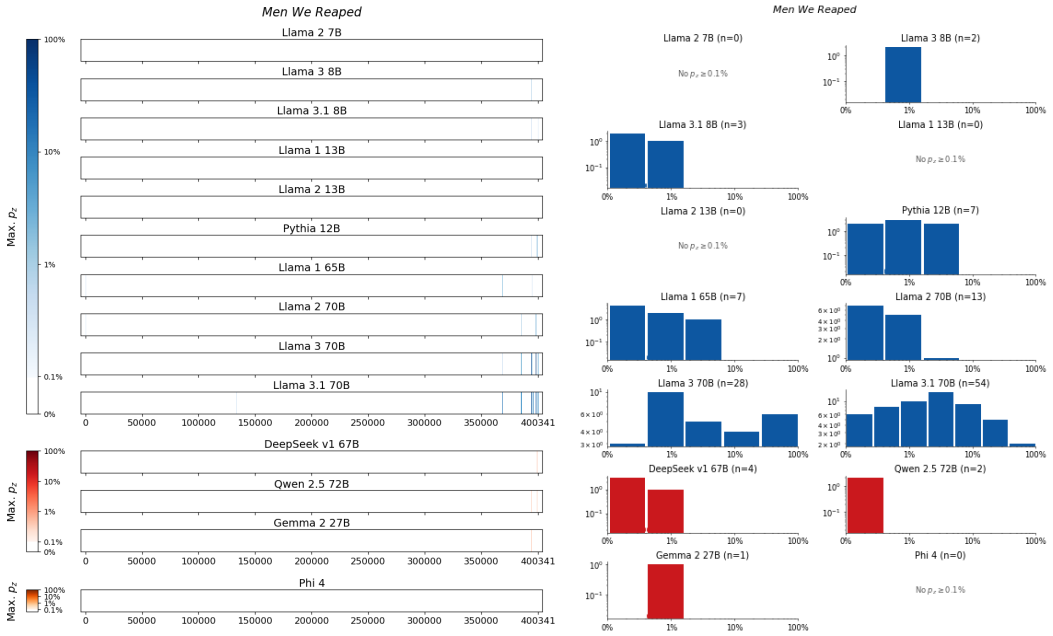


Figure 202: ***Men We Reaped,* Ward.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.192 *Charlotte's Web,* ***White***

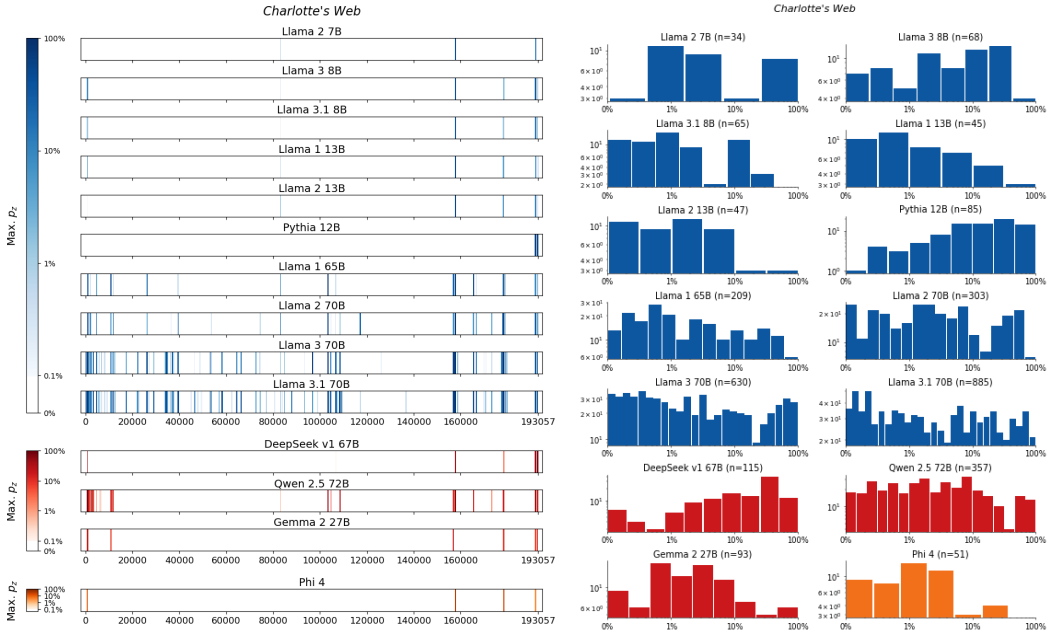


Figure 203: ***Charlotte's Web,* White.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.193 A Return to Love, Williamson

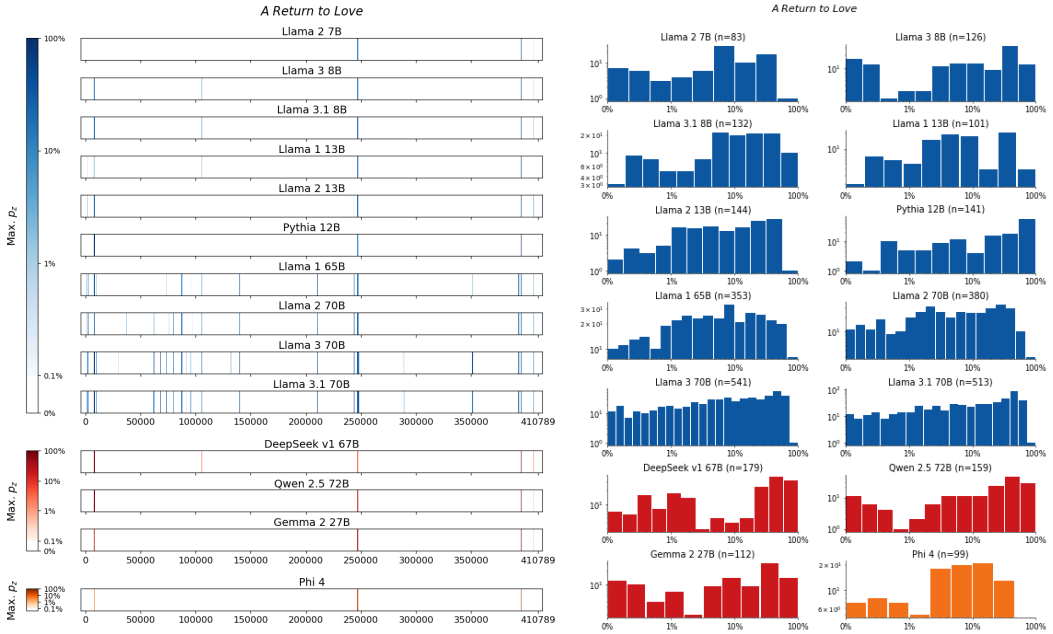


Figure 204: ***A Return to Love*, Williamson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.194 Another Brooklyn, Woodson

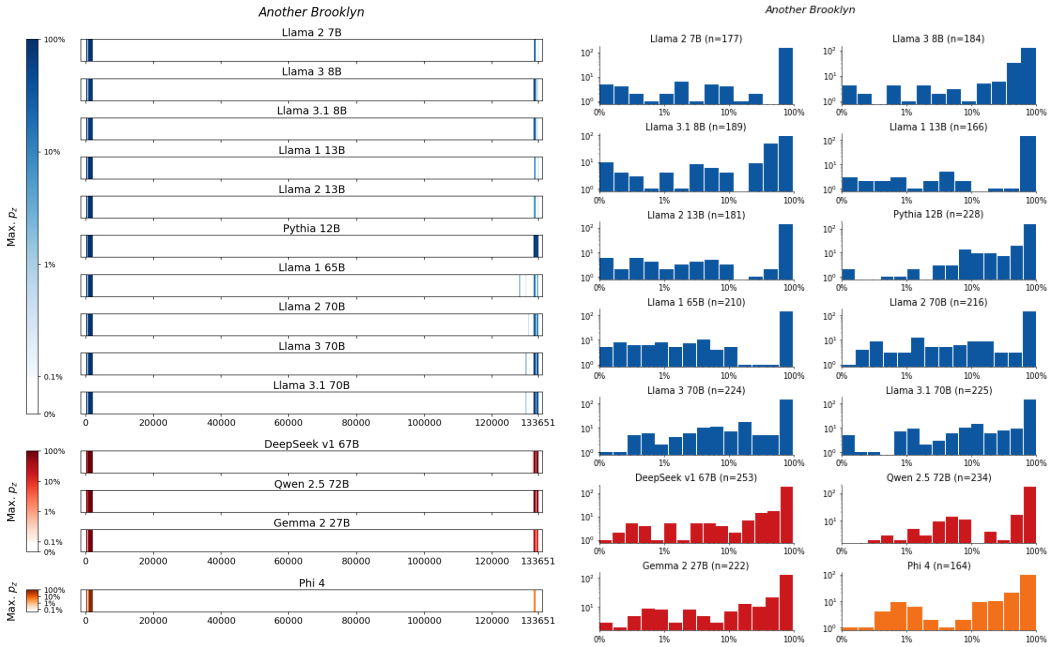


Figure 205: ***Another Brooklyn*, Woodson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.195 *Brown Girl Dreaming, Woodson*

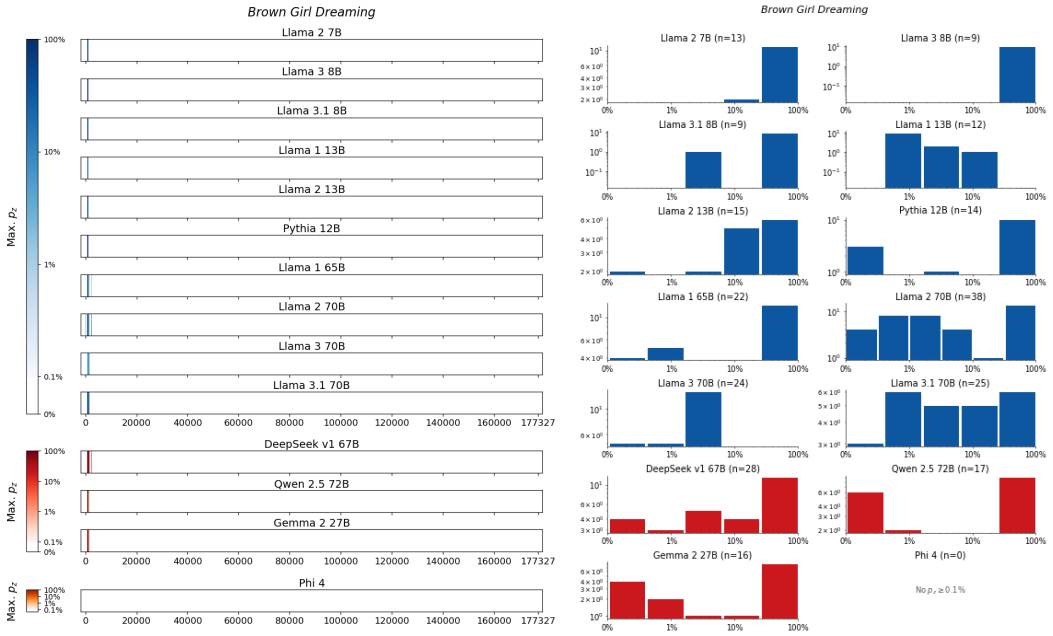


Figure 206: ***Brown Girl Dreaming*, Woodson.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.196 *A Little Life, Yanagihara*

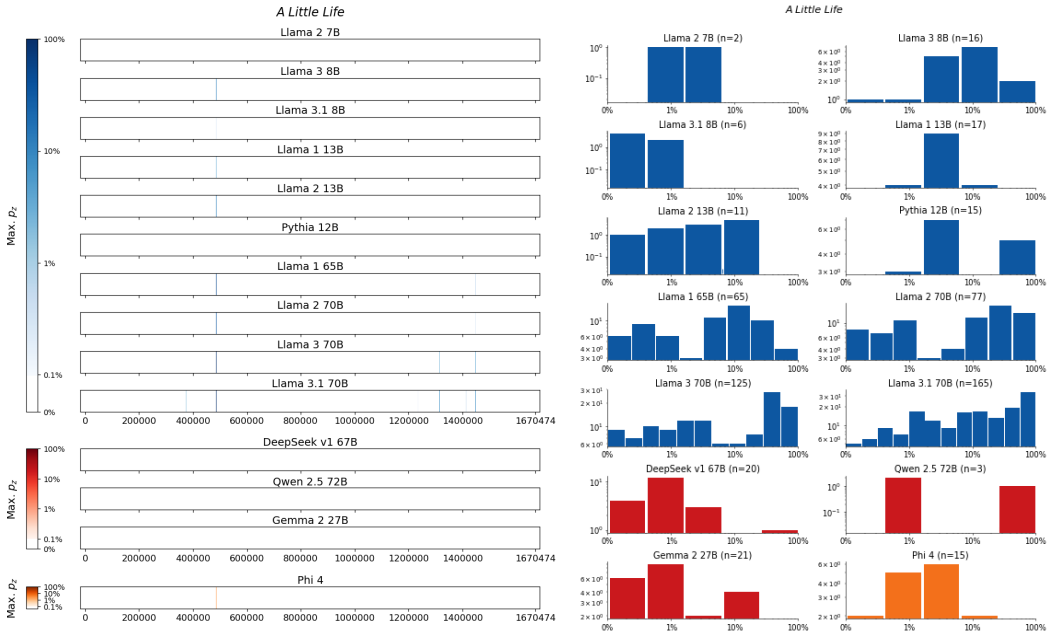


Figure 207: ***A Little Life*, Yanagihara.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.197 *The Art of Bonsai, Yoshimura & Halford*

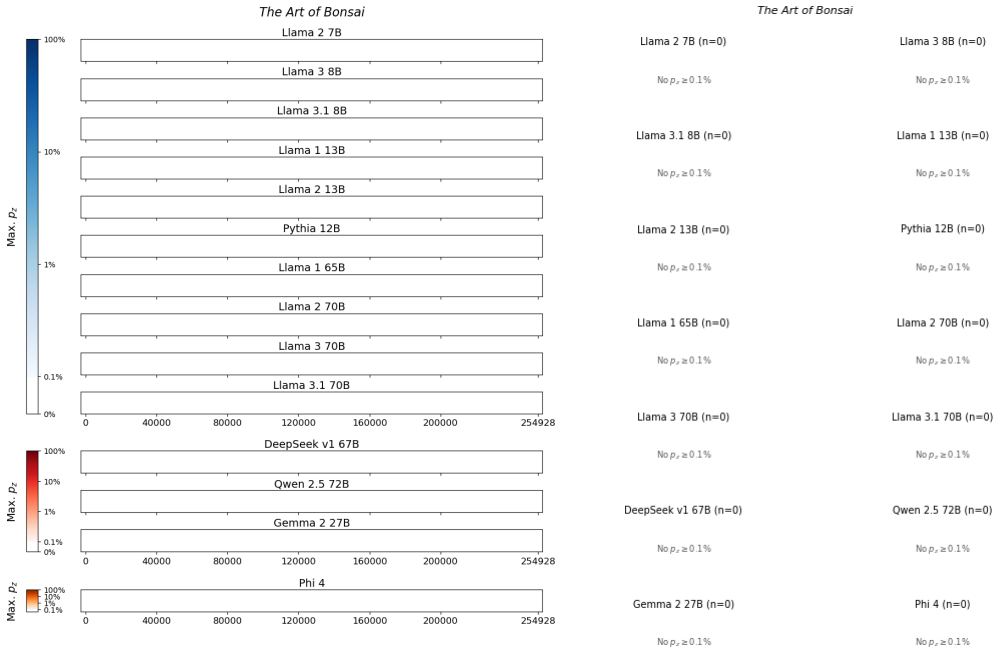


Figure 208: ***The Art of Bonsai*, Yoshimura & Halford.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.198 *A People's History of the United States, Zinn*

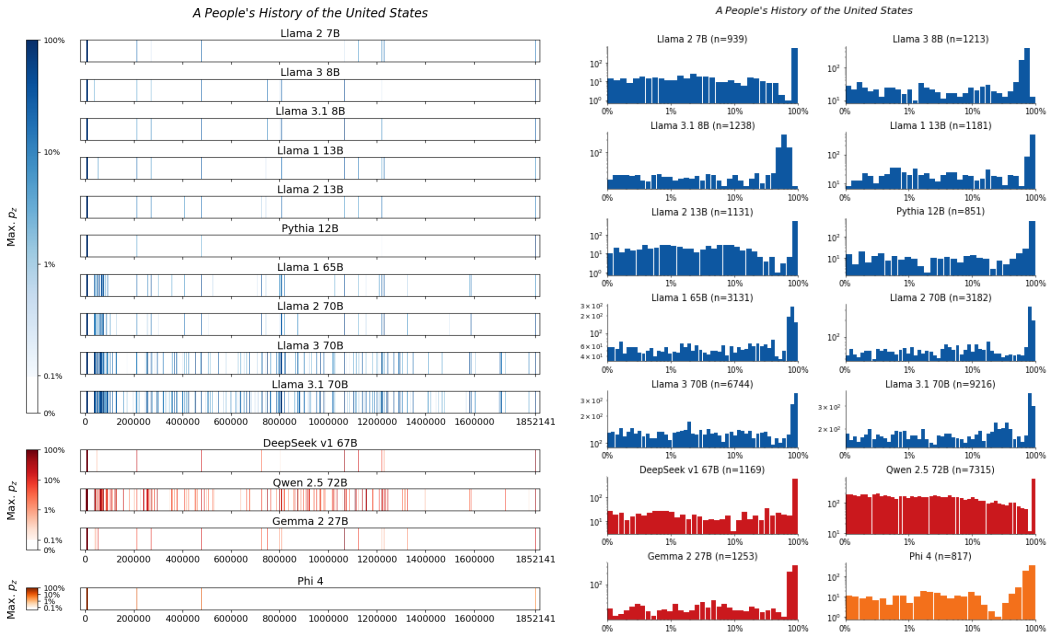


Figure 209: ***A People's History of the United States*, Zinn.** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.199 *The Future of the Internet and How to Stop It, Zittrain*

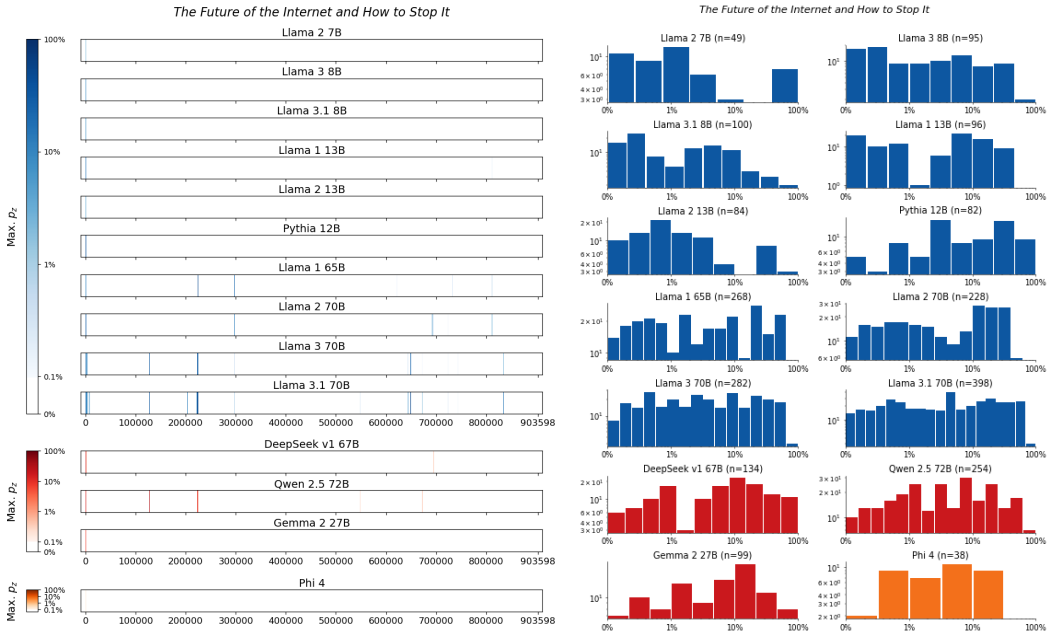


Figure 210: ***The Future of the Internet and How to Stop It, Zittrain.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.3.200 *The Book Thief, Zusak*

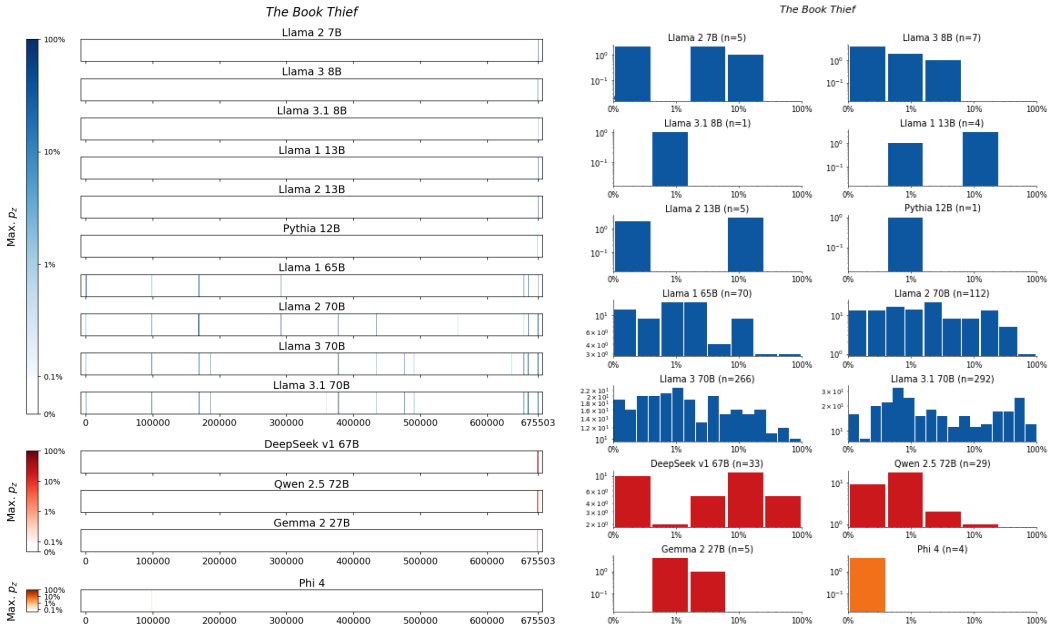


Figure 211: ***The Book Thief, Zusak.*** For 14 LLMs, (**left**) heatmaps for the sliding-window procedure and (**right**) corresponding distributions over suffix extraction probabilities ($\tau_{\text{min}} = 0.1\%$).

E.4 Estimating how much of a book is memorized by a model

We estimate book-specific memorization for a given model θ and decoding scheme ϕ with a metric for **extraction coverage**. We define this metric and provide measurements for the different models and books we evaluate.

E.4.1 Quantifying extraction coverage

We compute extraction coverage as the fraction of characters in a book that lie within at least one suffix whose extraction probability p_z exceeds a threshold τ.

Definition E.1 (Extraction coverage). Let $\mathsf{span}(z)$ denote the set of character indices spanned by a 50-token suffix of a given sequence $z \in \mathbb{Z}(B)$, where B is a book. Also let τ_{min} denote the minimum probability we consider to reflect extraction success. For book B, LLM θ, decoding scheme ϕ, and probability threshold $\tau \in [\tau_{\text{min}}, 1]$,

$$\mathsf{coverage}_{B,\theta,\phi}(\tau) = \frac{1}{L_B} \left| \bigcup_{z \in \mathbb{Z}(B):\, p_{z,\theta,\phi} \geq \tau} \mathsf{span}(z) \right|, \tag{8}$$

where L_B is the total number of characters in the book.

For details on τ_{min}, see Equation 7 (Sections 2 & 4; Appendices D & F). In the main text, we omit θ and ϕ for brevity, since we always specify the exact model and decoding scheme being considered.

The results of our sliding-window procedure can be used to compute Equation 8. This procedure returns a list of p_z sorted by start position in the book (i.e., $\mathbb{Z}(B)$). We pick a threshold τ and filter the list of p_z to only contain sequences whose extraction probability is $\geq \tau$. Our sliding-window procedure keeps track of character-level metadata so that we can visualize heatmaps; we have information about the start/end character indexes for the prefixes and suffixes. Based on these intervals, we can compute which characters in the whole book are covered with $p_z \geq \tau$, and then divide that by the total number of characters in the book.

High extraction coverage does not necessarily mean we can extract the whole book in one go, or that we have identified the minimal set of sequences to try to extract to get maximal coverage of the book. This is simply a procedure for estimating the proportion of total memorized text for a book, in order to do cross-book and cross-model comparisons, as in Section 3.2.

E.4.2 Extended extraction coverage results

We display extraction coverage in two ways, which support different types of comparisons.

1. **Model-level histograms.** For each model, we plot the distribution of extraction coverage across the 200 evaluated books. This provides a high-level sense of how coverage varies across books, and enables comparisons of this distribution across models. For example, in the main paper we show one histogram for LLAMA 3.1 70B at $\tau = 0.1\%$, the minimum extraction probability that we validate in this work. These results indicate that most books have little extraction coverage. But there are clear exceptions. They also reveal that some models exhibit substantially higher coverage than others—in this case, LLAMA 3.1 70B shows higher maximum coverage and more books with significant coverage, i.e., it has memorized much more book text than LLAMA 1 65B.
2. **Book-level percentages.** For each book, we compare extraction coverage across different models. This not only indicates how memorized a given book is by a particular model, but also highlights variation in memorization of a given book across models.

Here, we provide more extensive extraction coverage results. We include both types of comparisons for the 14 models for which detailed sliding-window results are given in Appendix E.3. All results are shown for $\tau = \tau_{\text{min}} = 0.1\%$, which corresponds to the total (largest) amount of extraction coverage revealed by our experiments. Figure 212 plots histograms of extraction coverage across all 200 books for each model, and Table 4 reports per-book extraction coverage results for each model.

Table 4: **Extraction coverage for** 200 **books and** 14 **LLMs.** We use Definition E.1 to compute extraction coverage for the results in Appendix E.3, which use 100-token sequences (50-token prefix + 50-token suffix) and top-k decoding ($T = 1$, $k = 40$). We set $\tau = \tau_{\text{min}} = 0.1\%$ to compute the total amount of extraction coverage that we observe.

Proportion of each book that is extractable with $p_z \geq 0.1\%$ for different models

Book	Llama 2 7B	Llama 3 8B	Llama 3.1 8B	Llama 1 13B	Llama 2 13B	Pythia 12B	Llama 1 65B	Llama 2 70B	Llama 3 70B	Llama 3.1 70B	DS v1 67B	Qwen 2.5 72B	Gemma 2 27B	Phi 4
A Game of Thrones	0.65%	0.75%	0.78%	0.76%	0.78%	0.67%	3.34%	3.07%	37.39%	39.39%	0.81%	1.25%	0.87%	0.59%
A Little Life	0.02%	0.03%	0.03%	0.05%	0.05%	0.04%	0.16%	0.16%	0.35%	0.46%	0.04%	0.03%	0.07%	0.04%
A People's History of the United States	1.30%	2.01%	2.05%	1.91%	1.66%	0.92%	6.86%	6.77%	17.91%	26.27%	1.71%	18.82%	2.05%	1.01%
A Perfectly Good Family	0.17%	0.24%	0.24%	0.14%	0.18%	0.27%	0.21%	0.19%	0.42%	0.58%	0.28%	0.28%	0.24%	0.22%
A Return to Love	0.64%	0.85%	0.85%	0.85%	0.94%	0.62%	2.76%	3.04%	4.35%	4.02%	1.08%	0.80%	0.73%	0.58%
A Visit from the Goon Squad	0.07%	0.00%	0.10%	0.07%	0.07%	0.20%	0.29%	0.36%	0.74%	1.06%	0.19%	0.22%	0.17%	0.11%
A Wrinkle in Time	1.44%	1.62%	1.57%	1.61%	1.69%	1.75%	4.54%	3.93%	10.00%	14.59%	2.32%	2.09%	1.57%	1.44%
After I'm Gone	0.19%	0.28%	0.28%	0.16%	0.27%	0.36%	0.39%	0.31%	0.40%	0.46%	0.41%	0.36%	0.28%	0.24%
Alice's Adventures in Wonderland	6.44%	18.39%	18.84%	18.36%	21.02%	1.65%	42.89%	42.72%	47.74%	48.80%	38.86%	38.00%	24.94%	8.33%
All the Onions	0.27%	1.61%	0.00%	0.27%	0.27%	2.61%	0.88%	1.19%	2.77%	2.77%	2.77%	2.55%	2.11%	0.00%
Along Came a Spider	0.38%	0.55%	0.55%	0.48%	0.46%	0.78%	0.71%	0.92%	1.10%	1.13%	1.13%	0.56%	0.41%	0.21%
Americanah	0.12%	0.13%	0.17%	0.14%	0.26%	0.17%	0.58%	0.64%	0.85%	0.98%	0.17%	0.26%	0.13%	0.14%
Anastasia on Her Own	0.00%	0.49%	0.40%	0.27%	0.30%	0.25%	0.57%	0.82%	0.79%	0.78%	0.25%	0.75%	0.25%	0.50%
And Then There Were None	1.44%	1.51%	1.91%	1.49%	1.82%	2.04%	2.44%	2.65%	4.69%	6.03%	3.45%	3.18%	1.92%	1.74%
Animal Rights	0.54%	0.61%	0.69%	0.63%	0.65%	0.59%	1.10%	1.00%	1.34%	1.46%	0.79%	0.97%	0.80%	0.48%
Another Brooklyn	2.04%	2.45%	2.41%	1.97%	2.17%	2.59%	3.01%	2.68%	2.84%	3.16%	2.85%	2.62%	2.46%	1.87%
Another Country	0.12%	0.16%	0.16%	0.13%	0.13%	0.18%	0.29%	0.34%	0.55%	0.74%	0.17%	0.22%	0.14%	0.10%
Ararat	0.37%	0.29%	0.29%	0.37%	0.37%	0.46%	0.43%	0.49%	0.43%	0.45%	0.56%	0.56%	0.42%	0.18%
Automating the News	1.26%	0.50%	0.56%	1.42%	1.55%	0.87%	1.86%	1.96%	1.02%	1.16%	1.55%	0.88%	1.15%	0.50%
Backroads Boss Lady	0.35%	0.31%	0.31%	0.35%	0.32%	0.43%	0.36%	0.39%	0.32%	0.32%	0.42%	0.43%	0.40%	0.00%
Bad Kid	0.78%	0.83%	0.87%	0.77%	0.82%	0.93%	0.87%	0.88%	0.91%	0.91%	0.94%	0.94%	0.89%	0.82%
Beloved	0.17%	0.69%	0.45%	0.35%	0.61%	0.09%	1.87%	3.58%	9.78%	12.54%	0.34%	1.55%	0.69%	0.32%
Billionaire Democracy	0.64%	0.50%	0.43%	1.19%	1.06%	0.34%	2.04%	2.11%	1.68%	1.69%	1.16%	0.80%	0.84%	0.36%
Birding with Yeats	0.06%	0.21%	0.20%	0.06%	0.06%	0.48%	0.33%	0.36%	0.62%	0.53%	0.51%	0.48%	0.40%	0.14%
Blink	0.31%	0.87%	0.87%	0.60%	0.63%	0.48%	2.19%	2.87%	6.86%	9.73%	1.51%	2.62%	0.87%	0.65%
Born to Walk	0.23%	0.17%	0.17%	0.27%	0.27%	0.16%	0.73%	0.76%	0.89%	0.86%	0.42%	0.61%	0.19%	0.10%
British Destroyers	0.03%	0.02%	0.00%	0.04%	0.05%	0.00%	0.08%	0.07%	0.06%	0.06%	0.08%	0.05%	0.02%	0.00%
Brown Girl Dreaming	0.31%	0.34%	0.34%	0.30%	0.32%	0.43%	0.81%	0.83%	0.54%	0.55%	0.80%	0.53%	0.58%	0.00%
Building and Operating a Realistic Model Railway	0.11%	0.00%	0.00%	0.11%	0.00%	0.20%	0.00%	0.11%	0.14%	0.14%	0.31%	0.22%	0.12%	0.00%
Buzz	0.05%	0.14%	0.20%	0.10%	0.10%	0.07%	0.58%	0.61%	2.16%	1.98%	0.24%	1.35%	0.42%	0.08%
Call Me Brooklyn	0.08%	0.09%	0.09%	0.11%	0.12%	0.11%	1.35%	0.79%	1.38%	1.59%	0.95%	0.09%	0.09%	0.08%
Catch-22	0.34%	0.57%	0.61%	0.60%	0.75%	0.32%	2.60%	2.98%	10.01%	15.64%	0.60%	0.93%	0.56%	0.23%
Catching the Sky	0.11%	0.20%	0.12%	0.20%	0.27%	0.37%	0.27%	0.27%	0.21%	0.21%	0.37%	0.30%	0.20%	0.00%
Charlie and the Chocolate Factory	0.82%	0.92%	0.92%	0.82%	1.00%	0.92%	4.49%	5.70%	17.09%	27.67%	1.50%	2.43%	3.83%	0.92%
Charlotte's Web	0.66%	1.71%	1.66%	1.47%	1.35%	0.82%	5.11%	7.23%	16.72%	21.96%	1.85%	6.47%	1.94%	1.29%
Chesapeake Requiem	0.58%	0.58%	0.58%	0.60%	0.64%	0.70%	0.68%	0.70%	0.70%	0.70%	0.71%	0.70%	0.68%	0.58%
Coal Creek	0.20%	0.00%	0.15%	0.25%	0.25%	0.33%	0.32%	0.22%	0.21%	0.29%	0.34%	0.28%	0.25%	0.13%
Competitive Strategy	0.09%	0.16%	0.12%	0.11%	0.11%	0.18%	0.25%	0.48%	1.70%	1.87%	0.21%	0.25%	0.26%	0.00%
Cosmos	0.33%	1.10%	1.39%	0.99%	0.90%	0.18%	3.64%	4.22%	7.73%	11.57%	0.63%	3.95%	0.99%	0.32%
Dante and the Origins of Italian Literary Culture	0.00%	0.04%	0.04%	0.00%	0.00%	0.04%	0.00%	0.00%	0.36%	0.55%	0.06%	0.08%	0.06%	0.08%
Dead Ringers	0.50%	0.42%	0.42%	0.47%	0.49%	0.66%	0.68%	0.56%	0.64%	0.65%	0.68%	0.69%	0.50%	0.28%
Dead Wake	0.07%	0.08%	0.12%	0.07%	0.10%	0.06%	0.13%	0.17%	0.47%	0.41%	0.18%	0.19%	0.06%	0.12%

Continued on next page

Table 4 (continued). Proportion of each book extractable with $p_z \geq 0.1\%$ for different models.

Book	Llama 2 7B	Llama 3 8B	Llama 3.1 8B	Llama 1 13B	Llama 2 13B	Pythia 12B	Llama 1 65B	Llama 2 70B	Llama 3 70B	Llama 3.1 70B	DS v1 67B	Qwen 2.5 72B	Gemma 2 27B	Phi 4
Down and Out in the Magic Kingdom	0.30%	0.21%	0.21%	0.31%	0.31%	0.47%	0.46%	0.56%	0.94%	1.11%	0.47%	0.48%	0.37%	0.00%
Drown	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%	0.37%	0.37%	0.12%	0.00%	0.00%	0.00%
Dungeons and Dragons and Philosophy	0.32%	0.31%	0.26%	0.45%	0.63%	0.86%	1.43%	1.31%	1.61%	1.62%	1.29%	0.91%	0.48%	0.31%
Eat More Plants	0.74%	0.69%	0.78%	0.74%	0.75%	0.73%	0.96%	1.26%	1.57%	1.61%	1.11%	1.49%	0.95%	0.70%
Embraced	5.25%	4.61%	4.81%	6.38%	5.40%	4.12%	6.92%	6.97%	5.68%	6.04%	6.70%	5.48%	4.76%	2.94%
Enlightenment Now	0.84%	1.19%	1.24%	1.05%	1.04%	0.59%	2.10%	2.16%	5.46%	5.16%	1.27%	2.21%	1.11%	0.84%
Essential Pepin Desserts	0.94%	1.10%	1.10%	0.88%	0.83%	0.91%	0.83%	0.88%	1.23%	1.34%	1.32%	1.41%	0.72%	0.98%
Ethnography after Antiquity	0.05%	0.05%	0.05%	0.05%	0.08%	0.13%	0.05%	0.21%	0.52%	0.69%	0.18%	0.24%	0.10%	0.00%
Fantastic Voyage	0.00%	0.00%	0.00%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Fear of Music	0.00%	0.33%	0.28%	0.00%	0.00%	0.40%	0.72%	0.80%	0.95%	0.96%	0.44%	0.33%	0.18%	0.00%
Fifty Shades of Grey	0.17%	0.14%	0.13%	0.12%	0.12%	0.15%	1.39%	1.03%	2.00%	2.71%	0.21%	0.16%	0.15%	0.13%
Fight Club	0.16%	0.27%	0.27%	0.41%	0.47%	0.26%	2.15%	3.58%	6.11%	6.65%	0.39%	0.26%	0.58%	0.27%
Florals & Botanicals	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Frankenstein	3.91%	15.87%	11.86%	11.32%	19.58%	0.64%	65.15%	65.67%	83.34%	83.52%	28.78%	35.26%	49.19%	21.92%
Girl in Translation	0.41%	0.27%	0.33%	0.37%	0.45%	0.49%	0.47%	0.46%	0.54%	0.59%	0.48%	0.48%	0.30%	0.00%
Gone Girl	0.19%	0.24%	0.24%	0.20%	0.38%	0.17%	0.81%	1.05%	1.69%	2.12%	0.29%	0.29%	0.25%	0.18%
Good Omens	0.10%	0.13%	0.09%	0.13%	0.20%	0.22%	1.50%	1.77%	3.98%	4.51%	0.21%	0.29%	0.20%	0.06%
Graft	0.05%	0.00%	0.00%	0.05%	0.05%	0.19%	0.05%	0.11%	0.11%	0.11%	0.22%	0.04%	0.04%	0.00%
Great Hair Days	0.68%	0.70%	0.70%	0.68%	0.68%	0.78%	0.68%	0.68%	0.70%	0.70%	0.77%	0.75%	0.71%	0.69%
Guam: Past and Present	0.06%	0.00%	0.00%	0.06%	0.06%	0.00%	0.06%	0.24%	0.11%	0.21%	0.00%	0.08%	0.08%	0.00%
Harry Potter and the Chamber of Secrets	0.23%	0.96%	0.65%	0.50%	0.67%	0.30%	12.99%	16.64%	92.44%	94.94%	2.14%	6.53%	2.67%	0.19%
Harry Potter and the Deathly Hallows	0.17%	0.43%	0.51%	0.42%	0.29%	0.04%	13.59%	9.29%	53.25%	64.79%	1.11%	4.06%	0.72%	0.13%
Harry Potter and the Goblet of Fire	0.15%	0.26%	0.39%	0.32%	0.39%	0.13%	26.15%	14.08%	82.35%	88.66%	1.09%	3.64%	1.12%	0.16%
Harry Potter and the Sorcerer's Stone	0.49%	2.15%	2.62%	1.97%	3.02%	0.41%	34.83%	36.41%	93.74%	96.27%	12.00%	58.21%	21.46%	1.73%
Industrial Magic	0.00%	0.06%	0.06%	0.00%	0.00%	0.06%	0.11%	0.06%	0.12%	0.07%	0.08%	0.06%	0.06%	0.00%
Inhuman Land	0.00%	0.00%	0.00%	0.02%	0.00%	0.03%	0.02%	0.02%	0.00%	0.00%	0.03%	0.03%	0.00%	0.00%
Invisible Man	0.80%	0.86%	0.92%	0.82%	0.78%	0.70%	1.02%	1.38%	2.89%	3.22%	0.88%	1.14%	0.84%	0.77%
James and the Giant Peach	1.16%	1.21%	1.21%	1.29%	1.16%	1.24%	2.03%	1.70%	3.26%	3.35%	1.79%	1.24%	1.64%	0.79%
Jesse James	0.09%	0.12%	0.23%	0.16%	0.21%	0.04%	0.86%	0.76%	0.74%	0.93%	0.48%	0.40%	0.08%	0.06%
Jurassic Park	0.00%	0.07%	0.07%	0.06%	0.00%	0.08%	0.51%	0.94%	4.03%	3.96%	0.08%	0.09%	0.08%	0.00%
Kitchen Table Wisdom	0.31%	0.34%	0.35%	0.41%	0.32%	0.23%	1.39%	1.70%	2.20%	2.95%	0.24%	0.51%	0.39%	0.00%
Knowing Your Value	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Lean In	0.70%	1.99%	1.71%	1.05%	1.48%	0.38%	4.61%	5.16%	8.76%	10.73%	2.49%	3.10%	1.49%	0.94%
Life on Air	0.13%	0.09%	0.11%	0.13%	0.15%	0.11%	0.23%	0.23%	0.23%	0.25%	0.17%	0.15%	0.12%	0.12%
Live and Learn	0.04%	0.00%	0.00%	0.04%	0.05%	0.13%	0.05%	0.04%	0.09%	0.09%	0.17%	0.17%	0.08%	0.00%
London in Chains	0.16%	0.12%	0.13%	0.16%	0.16%	0.22%	0.16%	0.16%	0.12%	0.13%	0.22%	0.20%	0.13%	0.12%
Lord of the Flies	0.88%	1.43%	1.43%	1.37%	1.29%	0.80%	4.85%	7.31%	29.10%	36.27%	2.12%	8.14%	10.05%	1.42%
Lullaby Town	0.26%	0.31%	0.30%	0.24%	0.27%	0.37%	0.33%	0.34%	0.38%	0.38%	0.35%	0.31%	0.29%	0.19%
M. Butterfly	0.91%	1.63%	1.63%	1.09%	1.13%	1.60%	1.46%	1.81%	2.46%	2.91%	1.72%	1.47%	1.20%	0.66%
Mark Rothko	0.10%	0.18%	0.17%	0.36%	0.20%	0.19%	0.78%	0.81%	1.32%	1.30%	0.47%	0.46%	0.12%	0.32%
Marvel's Spider-Man: Hostile Takeover	0.32%	0.32%	0.41%	0.32%	0.41%	0.42%	0.57%	0.67%	0.64%	0.85%	0.48%	0.44%	0.47%	0.23%
Men We Reaped	0.00%	0.06%	0.16%	0.00%	0.00%	0.17%	0.28%	0.33%	0.85%	1.13%	0.12%	0.15%	0.04%	0.00%
Middle India	0.38%	0.47%	0.47%	0.31%	0.33%	0.49%	0.39%	0.39%	0.50%	0.50%	0.50%	0.50%	0.45%	0.21%
Migrations to Solitude	0.15%	0.41%	0.42%	0.40%	0.39%	0.16%	0.98%	0.76%	1.23%	1.48%	0.59%	0.82%	0.58%	0.23%
Murder on the Orient Express	0.61%	0.78%	0.87%	0.53%	0.86%	1.88%	1.21%	1.48%	2.56%	2.48%	2.34%	1.79%	0.95%	0.78%
My Einstein	0.11%	0.35%	0.48%	0.24%	0.24%	0.00%	0.76%	0.73%	0.97%	1.01%	0.35%	0.70%	0.61%	0.00%

Continued on next page

Table 4 (continued). Proportion of each book extractable with $p_z \geq 0.1\%$ for different models.

Book	Llama 2 7B	Llama 3 8B	Llama 3.1 8B	Llama 1 13B	Llama 2 13B	Pythia 12B	Llama 1 65B	Llama 2 70B	Llama 3 70B	Llama 3.1 70B	DS v1 67B	Qwen 2.5 72B	Gemma 2 27B	Phi 4
Night Watch	0.14%	0.30%	0.30%	0.13%	0.23%	0.29%	0.37%	0.55%	1.55%	1.90%	0.35%	0.37%	0.29%	0.27%
Nineteen Eighty-Four	7.24%	13.27%	16.21%	16.77%	22.50%	1.88%	67.27%	71.21%	79.97%	81.19%	12.50%	30.59%	19.16%	5.09%
Nixon in China	0.15%	0.18%	0.25%	0.16%	0.17%	0.12%	0.77%	0.82%	1.17%	1.28%	0.25%	0.84%	0.18%	0.06%
No Visible Bruises	0.40%	0.49%	0.50%	0.44%	0.47%	0.65%	0.40%	0.56%	0.56%	0.56%	0.66%	0.59%	0.49%	0.49%
Noodles Every Day	0.07%	0.19%	0.26%	0.06%	0.07%	0.32%	0.07%	0.21%	0.33%	0.33%	0.34%	0.23%	0.08%	0.09%
Norwegian Wood	0.26%	0.19%	0.20%	0.22%	0.31%	0.33%	0.74%	0.72%	2.18%	2.52%	0.33%	0.27%	0.26%	0.11%
Notes of a Native Son	0.27%	0.93%	0.83%	0.81%	0.83%	0.16%	3.16%	3.73%	15.59%	16.80%	0.43%	1.06%	0.30%	0.32%
On Liberty	4.71%	21.65%	23.13%	11.10%	13.09%	2.42%	73.12%	49.66%	90.03%	90.83%	9.45%	80.53%	17.30%	7.93%
On the Road with Bob Dylan	0.00%	0.00%	0.00%	0.00%	0.00%	0.07%	0.00%	0.00%	0.00%	0.05%	0.07%	0.05%	0.00%	0.00%
Pagans	0.34%	0.40%	0.45%	0.33%	0.38%	0.48%	0.48%	0.49%	0.46%	0.47%	0.47%	0.50%	0.47%	0.32%
Paradise Valley	0.22%	0.35%	0.35%	0.11%	0.21%	0.42%	0.22%	0.22%	0.37%	0.37%	0.38%	0.41%	0.24%	0.21%
Payard Cookies	0.00%	0.00%	0.00%	0.00%	0.00%	0.15%	0.12%	0.10%	0.15%	0.15%	0.14%	0.15%	0.14%	0.00%
Pearl Harbor	0.87%	1.02%	1.10%	0.91%	0.85%	1.31%	0.91%	1.10%	1.17%	1.19%	1.22%	1.26%	1.04%	0.00%
Polaris	0.27%	0.31%	0.14%	0.34%	0.34%	0.44%	0.35%	0.39%	0.51%	0.54%	0.44%	0.44%	0.38%	0.16%
Portugal's Guerrilla Wars in Africa	0.05%	0.05%	0.05%	0.04%	0.00%	0.05%	0.11%	0.16%	0.14%	0.20%	0.06%	0.05%	0.04%	0.00%
Pride and Prejudice	3.14%	30.49%	56.87%	11.38%	22.10%	0.94%	72.56%	73.96%	83.74%	84.20%	33.42%	27.32%	10.13%	5.59%
Requiem for the Sun	0.08%	0.10%	0.10%	0.12%	0.11%	0.14%	0.15%	0.22%	0.14%	0.14%	0.20%	0.16%	0.09%	0.10%
Rome and Jerusalem	0.12%	0.34%	0.41%	0.16%	0.12%	0.10%	0.79%	0.59%	2.30%	3.06%	0.21%	0.57%	0.13%	0.10%
Rough-Hewn Land	0.22%	0.28%	0.32%	0.35%	0.31%	0.28%	0.71%	0.59%	1.05%	1.28%	0.54%	0.70%	0.43%	0.21%
Sally Ride: America's First Woman in Space	0.08%	0.10%	0.10%	0.13%	0.13%	0.16%	0.18%	0.13%	0.16%	0.22%	0.18%	0.21%	0.09%	0.06%
Sandman Slim	0.19%	0.29%	0.29%	0.23%	0.27%	0.36%	0.32%	0.32%	0.38%	0.43%	0.37%	0.38%	0.34%	0.25%
Selected Poems of Langston Hughes	1.27%	4.24%	3.65%	1.36%	1.61%	1.63%	6.66%	8.37%	17.76%	21.38%	5.33%	11.31%	6.36%	6.02%
Simple Cakes	0.60%	0.63%	0.66%	0.62%	0.64%	0.65%	0.63%	0.86%	1.04%	1.45%	0.93%	0.67%	0.66%	0.65%
Slaughterhouse-Five	0.93%	2.47%	2.37%	2.09%	2.12%	0.15%	8.04%	8.99%	19.84%	29.60%	1.57%	3.62%	2.24%	1.79%
Soft in the Head	0.00%	0.25%	0.24%	0.21%	0.27%	0.43%	0.27%	0.39%	0.42%	0.42%	0.89%	0.16%	0.00%	0.00%
Songs in Ordinary Time	0.13%	0.10%	0.13%	0.13%	0.13%	0.15%	0.13%	0.13%	0.16%	0.17%	0.15%	0.14%	0.14%	0.10%
Sunburn	0.27%	0.30%	0.30%	0.27%	0.27%	0.34%	0.29%	0.45%	0.44%	0.43%	0.38%	0.33%	0.33%	0.29%
Sweater Surgery	0.00%	0.45%	0.45%	0.00%	0.00%	0.82%	0.00%	0.00%	0.79%	0.79%	0.78%	0.81%	0.79%	0.00%
Tai Chi for Depression	0.65%	0.65%	0.65%	0.68%	0.66%	0.67%	0.82%	0.66%	1.20%	1.44%	0.71%	0.67%	0.68%	0.66%
The 7 Habits of Highly Effective People	0.45%	0.98%	1.28%	0.99%	1.21%	0.23%	7.50%	7.85%	29.19%	36.31%	2.18%	3.84%	2.17%	0.30%
The Alchemist	1.68%	3.91%	2.92%	2.72%	2.93%	0.66%	11.90%	10.75%	40.69%	49.06%	3.85%	5.84%	3.97%	1.04%
The Art of Bonsai	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
The Baghdad Clock	0.00%	0.00%	0.00%	0.00%	0.00%	0.12%	0.07%	0.11%	0.18%	0.11%	0.09%	0.00%	0.10%	0.00%
The Beautiful Struggle	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
The Bedwetter	0.38%	0.42%	0.42%	0.50%	0.61%	0.57%	0.60%	0.61%	0.62%	0.63%	0.69%	0.48%	0.58%	0.39%
The Book Thief	0.05%	0.13%	0.06%	0.11%	0.05%	0.06%	0.96%	1.35%	2.20%	2.53%	0.22%	0.22%	0.07%	0.13%
The Brief Wondrous Life of Oscar Wao	0.35%	0.43%	0.34%	0.42%	0.45%	0.42%	0.70%	0.88%	1.52%	1.76%	0.51%	0.52%	0.34%	0.00%
The Cat's Pajamas	0.62%	0.74%	0.74%	0.69%	0.86%	0.72%	0.99%	0.94%	0.98%	0.96%	0.68%	0.94%	0.72%	0.77%
The Catcher in the Rye	1.53%	1.60%	2.02%	2.29%	2.77%	0.82%	11.57%	18.14%	31.74%	41.30%	2.23%	4.35%	2.73%	1.02%
The Christmas Train	0.18%	0.20%	0.20%	0.18%	0.18%	0.20%	0.25%	0.25%	0.37%	0.48%	0.27%	0.19%	0.19%	0.19%
The Chronicles of Narnia	0.45%	1.02%	1.01%	1.15%	1.03%	0.00%	7.59%	8.42%	32.90%	36.47%	1.07%	2.34%	0.74%	0.46%
The Complete Joy of Homebrewing	0.12%	0.18%	0.18%	0.11%	0.14%	0.20%	0.16%	0.17%	0.24%	0.25%	0.20%	0.20%	0.19%	0.17%
The Complete Robot	0.23%	0.28%	0.28%	0.26%	0.25%	0.20%	0.27%	0.49%	4.81%	6.75%	0.21%	0.34%	0.21%	0.26%
The Confessions of Max Tivoli	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%	0.00%	0.00%	0.10%	0.11%	0.13%	0.00%	0.07%	0.00%
The Cult of Loving Kindness	0.00%	0.00%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.20%	0.20%	0.11%	0.00%	0.00%	0.00%

Continued on next page

Table 4 (continued). Proportion of each book extractable with $p_z \geq 0.1\%$ for different models.

Book	Llama 2 7B	Llama 3 8B	Llama 3.1 8B	Llama 1 13B	Llama 2 13B	Pythia 12B	Llama 1 65B	Llama 2 70B	Llama 3 70B	Llama 3.1 70B	DS v1 67B	Qwen 2.5 72B	Gemma 2 27B	Phi 4
The Curious Incident of the Dog in the Night-Time	0.36%	0.43%	0.40%	0.46%	0.46%	0.33%	1.82%	2.36%	7.47%	9.91%	0.57%	0.87%	0.39%	0.26%
The Da Vinci Code	0.30%	0.44%	0.37%	0.40%	0.44%	0.27%	1.47%	2.06%	6.27%	10.14%	0.63%	0.84%	0.27%	0.27%
The Daughter of Odren	0.00%	0.00%	0.00%	0.00%	0.60%	0.85%	0.60%	0.68%	1.57%	1.59%	0.85%	0.82%	0.76%	0.00%
The DevOps Adoption Playbook	1.02%	1.05%	1.05%	1.20%	1.15%	0.72%	1.50%	1.51%	1.50%	1.75%	1.41%	1.73%	1.24%	1.11%
The Doomsday Prophecy	0.41%	0.46%	0.47%	0.41%	0.40%	0.53%	0.61%	0.63%	0.65%	0.66%	0.59%	0.60%	0.55%	0.38%
The Dragon Never Sleeps	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.07%	0.07%	0.06%	0.06%	0.08%	0.00%	0.00%	0.00%
The Duchess War	0.09%	0.09%	0.09%	0.09%	0.09%	0.17%	0.09%	0.09%	0.09%	0.09%	0.23%	0.09%	0.09%	0.09%
The Dude Abides	0.59%	1.15%	1.31%	0.68%	0.65%	0.94%	2.03%	2.01%	2.93%	3.26%	1.64%	1.63%	1.11%	0.33%
The Fault in Our Stars	0.84%	1.35%	1.04%	1.54%	1.57%	0.42%	7.16%	5.61%	11.04%	9.59%	1.18%	0.99%	1.03%	1.03%
The Fellowship of the Ring	0.68%	2.68%	2.29%	1.71%	1.43%	0.39%	19.20%	17.88%	71.07%	79.75%	2.07%	13.78%	2.74%	0.61%
The Fugitive	0.38%	0.42%	0.42%	0.36%	0.38%	0.46%	0.38%	0.38%	0.42%	0.42%	0.49%	0.52%	0.50%	0.18%
The Future of the Internet and How to Stop It	0.17%	0.37%	0.37%	0.46%	0.33%	0.23%	1.13%	1.27%	1.88%	2.18%	0.40%	0.93%	0.41%	0.22%
The Girl with the Dragon Tattoo	0.04%	0.05%	0.05%	0.00%	0.03%	0.04%	0.17%	0.05%	0.81%	1.16%	0.08%	0.04%	0.00%	0.00%
The Goldfinch	0.09%	0.13%	0.07%	0.13%	0.12%	0.14%	0.71%	0.91%	0.82%	1.08%	0.17%	0.15%	0.08%	0.07%
The Gondola Maker	0.15%	0.16%	0.16%	0.15%	0.15%	0.15%	0.27%	0.15%	0.30%	0.40%	0.15%	0.16%	0.15%	0.16%
The Grapes of Wrath	0.24%	0.65%	0.63%	0.48%	0.68%	0.14%	2.81%	3.06%	6.62%	7.68%	0.44%	2.70%	0.89%	0.47%
The Great Gatsby	5.15%	11.44%	12.51%	14.28%	12.51%	1.45%	73.60%	71.82%	85.29%	88.77%	23.69%	26.86%	33.64%	8.71%
The Handmaid's Tale	0.17%	0.38%	0.29%	0.30%	0.42%	0.31%	2.82%	3.78%	6.59%	9.57%	0.34%	0.92%	0.28%	0.29%
The Heretic Queen	0.10%	0.10%	0.10%	0.10%	0.10%	0.17%	0.10%	0.10%	0.10%	0.10%	0.22%	0.17%	0.10%	0.10%
The Hitchhiker's Guide to the Galaxy - Omnibus	0.67%	1.34%	1.20%	1.26%	1.17%	0.29%	5.14%	5.96%	18.05%	23.48%	2.39%	1.86%	1.33%	0.31%
The Hobbit	1.11%	1.65%	2.00%	2.37%	2.38%	0.07%	19.78%	15.03%	60.82%	68.69%	2.25%	10.59%	2.55%	1.38%
The Hours	0.19%	0.88%	0.71%	0.72%	0.64%	0.21%	1.73%	1.75%	3.49%	3.77%	0.70%	1.47%	0.77%	0.85%
The Hunger Games	0.23%	0.24%	0.24%	0.36%	0.37%	0.21%	2.75%	3.32%	11.09%	14.21%	0.54%	1.35%	1.45%	0.36%
The Infernal Machine	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%	0.00%	0.10%	0.11%	0.10%	0.12%	0.11%	0.00%	0.00%
The Infinity Link	0.06%	0.05%	0.05%	0.07%	0.07%	0.15%	0.10%	0.07%	0.10%	0.10%	0.15%	0.15%	0.07%	0.00%
The Land Before Avocado	0.67%	0.67%	0.78%	0.78%	0.67%	1.05%	0.90%	0.98%	1.18%	1.20%	1.14%	0.94%	0.64%	0.68%
The Lemon Table	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.08%	0.26%	0.27%	0.00%	0.00%	0.00%	0.00%
The Lonely Soldier	0.06%	0.00%	0.00%	0.10%	0.06%	0.06%	0.19%	0.25%	0.10%	0.19%	0.13%	0.09%	0.09%	0.00%
The Making of a Mediterranean Emirate	0.41%	0.42%	0.45%	0.41%	0.45%	0.48%	0.58%	0.68%	0.64%	0.63%	0.67%	0.59%	0.45%	0.30%
The Memory Police	0.74%	0.75%	0.75%	0.83%	0.75%	0.88%	0.83%	0.83%	0.77%	0.77%	0.84%	0.80%	0.76%	0.73%
The Myth of Sisyphus	1.45%	4.01%	3.38%	2.12%	2.66%	0.82%	5.94%	9.19%	20.22%	24.90%	2.20%	5.90%	3.15%	2.32%
The Night Children	0.21%	0.00%	0.00%	0.21%	0.21%	0.32%	0.31%	0.31%	0.00%	0.00%	0.32%	0.30%	0.00%	0.00%
The Old Man and the Sea	0.34%	1.28%	0.61%	0.70%	0.56%	0.63%	7.83%	14.99%	63.91%	72.67%	4.36%	33.91%	4.29%	1.89%
The Outsiders	0.97%	0.95%	1.08%	0.99%	1.10%	1.07%	2.99%	5.63%	9.33%	11.50%	1.20%	1.43%	3.95%	0.78%
The Parthenon	0.10%	0.17%	0.15%	0.10%	0.23%	0.13%	0.25%	0.28%	0.39%	0.39%	0.21%	0.19%	0.18%	0.16%
The President's Vampire	0.43%	0.44%	0.42%	0.48%	0.42%	0.59%	0.56%	0.56%	0.56%	0.58%	0.60%	0.58%	0.53%	0.11%
The Pretender	0.19%	0.07%	0.07%	0.19%	0.19%	0.26%	0.26%	0.26%	0.25%	0.26%	0.25%	0.25%	0.22%	0.00%
The Second Summoning	0.04%	0.12%	0.13%	0.04%	0.04%	0.14%	0.21%	0.29%	0.18%	0.18%	0.26%	0.14%	0.12%	0.00%
The Seductions of Darwin	0.04%	0.12%	0.12%	0.21%	0.15%	0.12%	0.61%	0.58%	1.62%	1.97%	0.90%	0.70%	0.48%	0.14%
The Servants of Twilight	0.33%	0.33%	0.37%	0.39%	0.38%	0.47%	0.52%	0.44%	0.54%	0.55%	0.47%	0.45%	0.29%	0.09%
The Slippery Year	0.00%	0.00%	0.00%	0.10%	0.10%	0.10%	0.18%	0.19%	0.11%	0.11%	0.22%	0.14%	0.16%	0.00%

Continued on next page

Table 4 (continued). Proportion of each book extractable with $p_z \geq 0.1\%$ for different models.

Book	Llama 2 7B	Llama 3 8B	Llama 3.1 8B	Llama 1 13B	Llama 2 13B	Pythia 12B	Llama 1 65B	Llama 2 70B	Llama 3 70B	Llama 3.1 70B	DS v1 67B	Qwen 2.5 72B	Gemma 2 27B	Phi 4
The Stone Sky	0.21%	0.31%	0.30%	0.38%	0.33%	0.25%	0.60%	0.79%	0.85%	1.19%	0.29%	0.37%	0.17%	0.09%
The Subtle Knife	0.07%	0.10%	0.10%	0.15%	0.00%	0.00%	0.38%	0.18%	2.21%	1.75%	0.08%	0.00%	0.07%	0.00%
The Third Man	0.97%	1.43%	1.08%	1.30%	1.19%	1.46%	1.86%	2.04%	2.02%	2.50%	1.62%	0.82%	0.79%	1.04%
The Tide Was Always High	0.00%	0.08%	0.09%	0.10%	0.10%	0.08%	0.16%	0.24%	0.33%	0.33%	0.14%	0.15%	0.13%	0.08%
The Water Dancer	0.63%	0.63%	0.63%	0.63%	0.69%	0.71%	0.68%	0.74%	0.70%	0.91%	0.76%	0.72%	0.69%	0.63%
The White Album	0.30%	0.31%	0.30%	0.31%	0.28%	0.16%	1.15%	1.61%	3.12%	3.67%	0.43%	0.31%	0.32%	0.00%
The World's Wife	0.73%	1.44%	1.44%	1.39%	0.85%	2.16%	3.20%	3.13%	7.41%	7.56%	2.08%	2.78%	2.24%	0.69%
Things Fall Apart	0.72%	1.31%	1.07%	0.63%	0.70%	0.18%	4.71%	6.84%	25.04%	31.99%	0.75%	13.56%	6.91%	1.69%
This Is How You Lose Her	0.66%	0.52%	0.52%	0.53%	0.66%	0.70%	0.86%	1.02%	1.11%	1.15%	0.69%	0.69%	0.41%	0.00%
Toscanini	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	0.00%	0.02%	0.02%	0.03%	0.02%	0.02%	0.00%
Tree and Leaf	1.69%	4.07%	4.73%	3.50%	4.67%	0.63%	12.29%	17.86%	35.63%	38.66%	3.24%	14.40%	3.56%	3.42%
Twilight	0.05%	0.00%	0.00%	0.00%	0.05%	0.00%	2.25%	1.65%	4.82%	5.67%	0.05%	0.05%	0.16%	0.00%
Ulysses	0.19%	0.81%	0.61%	0.27%	0.33%	0.10%	1.55%	10.14%	65.68%	76.11%	0.44%	0.88%	0.66%	0.36%
Uncommon Type	0.80%	0.74%	0.74%	0.80%	0.80%	0.93%	0.93%	0.95%	0.83%	0.83%	1.06%	0.96%	0.83%	0.75%
Unglued	3.58%	4.46%	4.43%	4.32%	4.21%	3.07%	4.84%	4.58%	4.96%	5.20%	4.85%	4.55%	3.49%	1.74%
Waiting for Godot	0.00%	1.63%	1.71%	1.66%	2.33%	1.41%	11.64%	13.83%	61.96%	71.15%	3.65%	6.67%	9.22%	0.41%
We Should All Be Mirandas	0.55%	0.55%	0.55%	0.55%	0.70%	0.79%	0.68%	0.70%	0.55%	0.56%	0.77%	0.80%	0.80%	0.56%
We Were Eight Years in Power	1.21%	1.37%	1.51%	1.59%	1.45%	0.89%	8.40%	7.92%	12.06%	14.26%	1.42%	4.40%	1.43%	0.74%
White Teeth	0.04%	0.05%	0.05%	0.05%	0.04%	0.05%	0.20%	0.41%	0.75%	0.89%	0.05%	0.09%	0.05%	0.05%
Who Is Rich?	0.85%	0.87%	0.87%	0.86%	0.85%	1.04%	0.93%	0.99%	0.98%	0.97%	1.04%	1.00%	0.95%	0.84%
Why New Orleans Matters	1.24%	1.36%	1.45%	1.22%	1.33%	1.54%	1.40%	1.43%	1.50%	1.50%	1.52%	1.52%	1.46%	1.19%
Windfall	1.93%	0.39%	0.30%	2.43%	2.30%	0.69%	3.33%	3.34%	1.04%	1.15%	2.40%	0.99%	1.59%	0.18%
Winnie the Pooh	2.21%	2.71%	3.69%	3.45%	3.64%	1.01%	26.14%	40.18%	77.31%	79.94%	4.86%	5.82%	3.96%	1.48%
Winter Sisters	0.95%	0.94%	0.94%	0.90%	0.95%	0.95%	1.01%	1.01%	0.95%	0.95%	1.02%	0.95%	0.94%	0.80%
Wizard's First Rule	0.08%	0.09%	0.09%	0.13%	0.10%	0.05%	0.18%	0.18%	0.25%	0.30%	0.13%	0.09%	0.12%	0.05%
Zombie Halloween	1.10%	0.59%	0.96%	1.05%	0.94%	1.16%	1.10%	1.18%	1.10%	1.11%	1.44%	0.98%	0.94%	0.25%

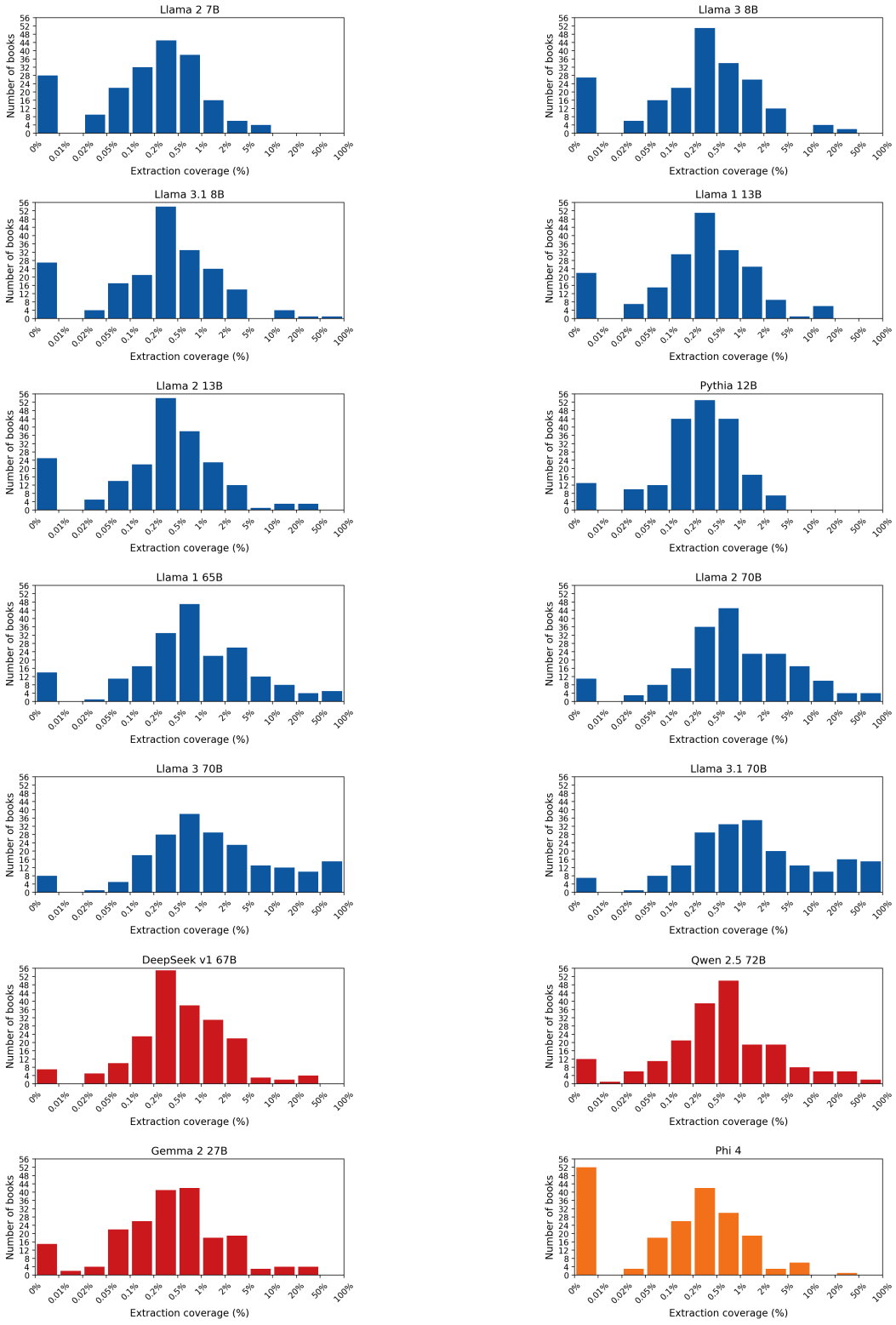


Figure 212: **Extraction coverage for** 200 **books and** 14 **LLMs.** Computing Definition E.1 for the results in Appendix E.3 (50-token prefix + 50-token suffix; top-k decoding with $T=1$, $k=40$; $\tau = \tau_{\text{min}} = 0.1\%$). We divide results into our three categories by color: LLMs that we know with certainty were trained on Books3 (blue), LLMs that we can confidently conclude were trained on (at least parts of) some books that were also contained in Books3 (red), and an LLM whose developers claim was not trained on whole copyrighted books (orange).

F Measurement validity

In the follow-on work of Cooper et al. (2026), we take up this topic in significantly more detail. Here, we provide more detail on the validity experiments introduced in Section 4. These experiments support that our measurements capture true instances of extraction. On this basis, we expand our analysis to LLMs with undisclosed information about training on books—DEEPSEEK V1 67B, QWEN 2.5 72B, and GEMMA 2 27B. Because our controls show that our method does not mistake chance generation for extraction, we can be confident that when these LLMs register extraction success with our measurements, it reflects memorization.

Given the length of this appendix, we provide a summary:

- **Extraction of known training data is evidence of memorization.** For models where the training data are known, high-probability generation of a verbatim suffix given its prefix is widely accepted as evidence of memorization (Carlini et al., 2021; 2023; Hayes et al., 2025b). Probabilistic extraction provides a more fine-grained version of the same test: rather than returning a single greedy continuation, it assigns a probability to the target suffix and reveals whether it surpasses a minimum acceptable threshold.
- **We adopt a deliberately conservative minimum threshold.** Probabilistic extraction produces a continuous measure of extraction, but requires a cutoff. We set τ_{min} very high to ensure we do not over-count memorization. This makes our results lower bounds: we under-count memorization (Section 2). We capture more instances of memorization than greedy decoding, but we still do not over-count. We defer more fine-grained settings of τ_{min} to future work.
- **Validating against non-training data** The majority of our experiments—those on LLAMA and PYTHIA models—fit the case of extracting known training data from models. It is *known* that `Books3` was included in the training data, and we are running experiments to extract `Books3` sequences. But we can also do more to ensure validity: we can test our procedure on known *non*-training data, and show that suffix probabilities fall below our chosen τ_{min}. We (and Hayes et al. (2025b)) do this for an additional sanity check. In particular, we perform over 300 negative control experiments (Section 4 & Appendix F.4). Work on greedy-decoded discoverable extraction often does not do this; it assumes validity or uses shorter prefix lengths to induce less conditioning on the output (Nasr et al., 2023) (Appendix F.6). Overall, our results confirm that our choice of τ_{min} is very conservative; we could set it lower without risking false positives. However, we are deliberately cautious, and defer this to future work.
- **Stress-testing extraction with longer prefixes.** When validating τ_{min} on known training data, we can also ask whether longer prefixes make memorization more discoverable. Because suffix probabilities are conditional on the prefix, extending the prompt can further condition the model's output toward a certain part of its learned distribution (which can also further be affected by the decoding algorithm). As a result (as others have shown (Carlini et al., 2023)), for experiments on *known* training data, when a piece of training data is memorized, a longer prefix prompt can raise the probability of the target suffix and expose it as memorized (via greedy decoding). This memorization may not be *discoverable* with only a 50-token prompt, but may become discoverable with a longer prompt. Our own experiments confirm these patterns, but also show that, for a specific sequence, that longer prompts do not always lead to successful extraction: if a sequence is not memorized, we are unable to find a long enough prefix to make it it extractable (Appendix F.3). This supports the conclusion that probabilistic extraction captures true memorization, and clarifies the limits of analogies that treat LLMs as "monkeys at the typewriter" (Appendix F.5).
- **Bringing the pieces together.** Our validity experiments combine baselines, negative controls, and varied-length prefix tests. For models known to be trained on `Books3`, we run baselines with different prefix lengths as points of comparison (Appendix F.3). We then add two sets of negative controls: (i) PHI 4, a model not trained on whole copyrighted books (Appendices E & F), and (ii) books published after the training cutoff for LLMs trained on `Books3` (Appendix F.4.2). Across more than 300 such experiments, suffix probabilities for non-training data are essentially zero, and nonzero values

are orders of magnitude below τ_{min}. As a result, our procedure does not mistakenly register non-training data as extractable, reinforcing that what we count as extraction reflects true instances of memorization.

- **Applying the procedure to models with unknown training data.** Based on these results, we can also apply the procedure to models whose training data are undisclosed. For sequences that register as extractable in these models, we can confidently interpret them as evidence of memorization, regardless of whether the training-data source was `Books3` directly or overlapping material from elsewhere (Appendix F.6). Importantly, as we state throughout, for our experiments on these models, our claims about membership in the training data *are only about specific extracted sequences*, not whole books.

Appendix outline.

- **Appendix F.1.** Motivation for these experiments.
- **Appendix F.2.** Setup.
- **Appendix F.3.** Additional baseline experiments for LLMs known to be trained on `Books3`.
- **Appendix F.4.** Additional negative control experiments on non-training data.
- **Appendix F.5.** Additional discussion on "the monkey at the typewriter."
- **Appendix F.6.** Discussion of relationship to prior literature on membership inference.

F.1 Motivating our validity experiments

The main point of these experiments is to validate that our measurements capture true instances of extraction, rather than chance generation of verbatim suffixes. They help us make confident claims about both our results for models trained on `Books3` and our application of the procedure to models with undisclosed training data. Here, we expand on these motivations (Appendix F.1.1), connect our work to the literature on greedy-decoded discoverable extraction (Appendix F.1.2), and clarify the role of varying prefix length in our baselines and negative controls (Appendix F.1.3).

F.1.1 Setting τ_{min} and extending to models with unknown training data

We can be fairly confident that DEEPSEEK V1 67B, QWEN 2.5 72B, and GEMMA 2 27B were trained on at least some books: the Pile (Gao et al., 2020) and several other common LLM pre-training datasets either directly include `Books3` or other book corpora (Lee et al., 2023a;b). Other work on extraction with uncertain training data treats the Pile as part of a proxy for the true training set (Nasr et al., 2023; 2025). However, without disclosure, we cannot be absolutely certain. And without knowing whether `Books3` was in fact included, we cannot—without additional validity experiments—make claims about memorization that are as strong as those we make for models known to be trained on `Books3`. This is because membership in the training data is a prerequisite for memorization; by definition, a model cannot memorize sequences it has never seen.

This is also why observing p_z on non-training data is important for setting τ_{min}. By comparing suffix probabilities from negative controls on non-training data (Section 4, Appendix F.4) with those from baselines on known training data (Section 4, Appendix F.3), we can draw a conservative line for what p_z is high enough to be indicative of memorization.

That is, our negative controls clarify how to set the threshold that we use to count extraction success, τ_{min}—the minimum p_z for which we interpret our measurements as evidence of memorization. Running our procedure on *non*-training data, we confirm that the values of p_z we observe are far below our chosen cutoff. That way, when we count extraction success for a given sequence with $p_z \geq \tau_{\text{min}}$, we can be confident that it does not reflect chance generation. Our results in Section 4 and the rest of this Appendix confirm this: $\tau_{\text{min}} = 0.1\%$ is conservative—well above any p_z observed in negative control experiments.

We therefore use these results to justify our choice of τ_{min} in this paper. More extensive experiments would be needed to lower τ_{min} or calibrate per LLM. Here, we prefer to be

conservative: we would rather under-count memorization than risk over-counting (Carlini et al., 2022; Hayes et al., 2025a), and defer fine-grained threshold settings to future work.

F.1.2 Relating greedy-decoded and probabilistic discoverable extraction

To our knowledge, we run the largest set of validity experiments in the ML literature on memorization and extraction. Such experiments are often not performed in work on extraction (Nasr et al., 2023; 2025; Carlini et al., 2021; Karamolegkou et al., 2023), with important exceptions (Carlini et al., 2023). Hayes et al. (2025b) conducted one small-scale test on non-training data for a small PYTHIA model, and used this to support validity claims for probabilistic extraction from much larger and higher quality LLAMA 1 models.

Given these norms, we initially omitted our negative controls from the main paper, deferring discussion to the Appendix. However, after receiving some feedback about earlier drafts of this work that indicated serious misunderstandings about the nature of memorization and extraction, we realized that including the results of these experiments in the main text could help some readers better understand our results and claims. We now clearly delineate two sets of negative controls, both in this appendix and in the main paper text (Section 4): the experiments on PHI 4 (Appendix F.4.1) and the experiments on books published after training-date cutoffs (Appendix F.4.2).

Here, we clarify the relationship between prior definitions and our approach, and explain how this motivates our validity experiments.

Greedy-decoded discoverable extraction. It is widely accepted that successful greedy-decoded discoverable extraction on known training data is evidence of memorization; memorization remains the most parsimonious explanation (Hayes et al., 2025b; Carlini et al., 2021; Nasr et al., 2023; Carlini, 2025; Cooper & Grimmelmann, 2024). In this procedure, greedy decoding always selects the top-ranked token at each step, making it a deterministic approximation of the most likely sequence. Carlini et al. (2021) describe this choice as a tractable proxy for the arg max in their definition of model knowledge extraction:

Definition F.1 (**Model Knowledge Extraction**, from Carlini et al. (2021))**.** A string s is extractable from an LM f_θ if there exists a prefix c such that:

$$s \leftarrow \underset{s':|s'|=N}{\arg\max}\, f_\theta(s'|c)$$

We abuse notation slightly here to denote by $f_\theta(s'|c)$ the likelihood of an entire sequence s'. *Since computing the most likely sequence s is intractable for large N*, the arg max in [the definition] can be replaced by an appropriate sampling strategy *(e.g., greedy sampling)* that reflects the way in which the model f_θ generates text in practical applications.

We add emphasis to two areas in this definition, which we discuss below. First, we note that this definition does not mention training data; it is just a definition about a model having knowledge of string s. Carlini et al. (2021) connect extraction *of training data* to memorization through a second definition:

Definition F.2 (**k-Eidetic Memorization**, from Carlini et al. (2021))**.** A string s is k-eidetic memorized (for $k \geq 1$) by an LM f_θ if s is extractable from f_θ and s appears in at most k examples in the training data $X : |\{x \in X : s \subseteq x\}| \leq k$.

Successful extraction, under these definitions, is evidence for memorization. Ideally, s should be computed as the *most likely sequence* under f_θ. However, since this is not tractable to compute in practice, *greedy decoding* is suggested as a reasonable proxy. It forces top-1 choices regardless of their actual probabilities under the model.

The relationship to probabilistic extraction. Greedy decoding is equivalent to top-k decoding with $k=1$, which only considers the token with top-1-ranked logit. (Temperature has no effect in this setting, since both temperature and top-k are rank-preserving.) Probabilistic extraction relaxes this by considering a wider set of high-probability continuations (e.g., top-40). This better reflects the model's high-probability distribution and avoids the

problem where greedy decoding yields globally low-probability sequences. Hayes et al. (2025b) motivate their metric precisely from this point: the greedy-decoded suffix is *locally* optimal, but it may lead to a *globally* relatively low probability sequence under the model.

A hypothetical example illustrates this gap: a greedy-decoded suffix can have vanishing total probability, while a slightly different path within the top-k yields a much higher-probability suffix. Consider a greedy-decoded suffix where the first token has 99% probability and is selected, and then every top-1 token after it for 49 tokens has probability 1%; the total probability for the 50-token suffix would be $0.99 \times (0.01)^{49} = 9.9 \times 10^{-99}$. Now imagine we instead were *not* restricted to always picking the top-1 probability token; we relax this requirement to, for example, allow to select from the top-k tokens. We consider probabilities computed with respect to the base $T = 1$ distribution. For the first token position, the top-2 token might have probability 1%, and then the remaining top-1 49 tokens could each have probability 99%; the total probability for this suffix would be $0.01 \times (0.99)^{49} \approx 0.006$ (or about 0.6%).

The point is, by allowing for a bit more flexibility in the decoding strategy, the model can potentially produce much higher probability sequences; when we observe that such ($\geq \tau_{\text{min}}$) sequences match the verbatim suffix from the training data, then we count them as extracted. As long as we choose τ_{min} appropriately—i.e., such that it does not capture low-probability sequences that are perhaps reflective of generalization—we should obtain higher fidelity measurements of memorization. We should have fewer false negatives than greedy-decoded discoverable extraction.

Greedy-decoding is a degenerate case of probabilistic extraction. Greedy decoding can be seen as a degenerate case of probabilistic extraction, corresponding to $\tau_{\text{min}} = 100\%$. Each step assigns all probability mass to a single token under a normalized top-1 distribution, even if that token's true model probability is very low. Probabilistic extraction surfaces this mismatch more clearly and is a better proxy for identifying high-probability memorized sequences. Because it includes the greedy suffix when it matches the training suffix, it never misses those cases, while also capturing others that greedy decoding fails to expose.

F.1.3 The role of prefix length in our experiments

We run our validity experiments with various prefix lengths, not just for 100-token sequences with 50-token prefixes and 50-token suffixes. This is a deliberate choice because additional context conditions the model's output distribution toward a particular output space. For memorized sequences, this is known to surface memorization that may not appear for shorter prefixes. Carlini et al. (2023) call this the **discoverability phenomenon**: "some memorization only becomes apparent ...when the model is prompted with a sufficiently long context." Through the lens of probabilistic extraction, a longer training-data prefix can raise the probability of the target suffix in Equation 1, such that it surpasses τ_{min}. (It is worth noting that any conditional extraction procedure—i.e., one that uses prompts, rather than unconditionally generating from the LLM—statistically biases the output. This is one of the reasons that extraction under-counts memorization; it only searches a part of the learned distribution.)

By testing various prefix lengths on known training data, we can estimate an upper bound on memorization (with respect to our extraction procedure) for a given work. We expect longer prefixes to expose more memorization, but that increasing prefix length will not typically induce all suffix probabilities to 100%; that is, we will observe more extraction only if there is more memorization to uncover. It is not the case that one can simply make the prefix arbitrarily long and then "extract" any target suffix. We discuss this further in Appendices F.3 and F.5. This logic also strengthens the interpretation of our negative controls. In this case, our measurements should not capture extraction for unique sequences of non-training data—even when we intentionally condition with longer prompts. Again, this shows that arbitrarily long suffixes do not all of a sudden mean we will register extraction with our measurements. For these experiments, we should not be able to do so because the sequences were not included in the training data.

F.2 Setup

Books3 books. Running experiments with longer sequences is significantly more expensive. We therefore limit our results on varied prefixes to a small selection of books from Books3: *Beloved* (Morrison, 1987a), *Harry Potter and the Sorcerer's Stone* (Rowling, 1997), *This Is How You Lose Her* (Díaz, 2012), and *We Were Eight Years in Power* (Coates, 2017). We picked these 4 books because each represents (roughly speaking) a different pattern for memorization in our sliding-window results on 50-token prefixes (Appendix E) for the model that exhibits the most memorization: LLAMA 3.1 70B. *This Is How You Lose Her* is barely memorized at all (Appendix E.3.52). *Beloved* has fragmented memorization "hot-spots;" there are many throughout the book, but they are rather disconnected, separated by lots of regions that do not register any memorization (Appendix E.3.128). *We Were Eight Years in Power* has several "hot-spots;" some are fragmented, but some indicate long, continuous regions of memorized text (Appendix E.3.33. *Harry Potter and the Sorcerer's Stone* is effectively memorized entirely, though not all suffixes have equally high extraction probability (Appendix E.3.153).

Non-training data. The large majority of sequences in the in-copyright Books3 books above should not be training data for PHI 4; exceptions are highly duplicated text, like boilerplate copyright notices and occasional popular quotes that are included from other sources (e.g., the Bible, newspaper articles, earlier books). We also purchased a set of 4 books published in 2025—after the release date of the newest model we test, QWEN 2.5 72B (September 19, 2024). (See Table 5.) We run these books through models known to have been trained on Books3 as another point of comparison. Similar to the above note about PHI 4, none of the unique text in these books should ping as extractable. This may not be true for non-unique text: boilerplate text and popular or otherwise duplicated quotes from earlier sources.

Table 5: **Books that are non-training data.** The 4 books that post-date training cutoff for all LLMs we test. We use these books for one set of our negative control experiments. See Section 4 and Appendix F.4.2.

	Author	Title	Publication Date	Status
1	Gareth Brown	*The Society of Unknowable Objects* (Brown, 2025)	July 31, 2025	©
2	Emily Henry	*Great Big Beautiful Life* (Henry, 2025)	April 22, 2025	©
3	Ocean Vuong	*The Emperor of Gladness* (Vuong, 2025)	May 13, 2025	©
4	Sarah Wynn-Williams	*Careless People* (Wynn-Williams, 2025)	March 11, 2025	©

Overall, we want to see how changing prefix length impacts how much more memorization we can discover. As we increase prefix length, we expect to discover more memorization of Books3 text for models trained on Books3, and to effectively observe no major changes for PHI 4 for unique text from these books nor for unique text from the 2025 books.

Models. Following from above, we run these experiments on 3 models known to have been trained on Books3. We include LLAMA 3.1 70B, because this is the model that has memorized the most of all LLAMA models we test. We also include LLAMA 3.1 8B, as it is a much smaller model from the same family. It memorizes less than LLAMA 3.1 70B. (See Appendix E.) While it may be possible that increasing prefix length for LLAMA 3.1 70B surfaces many more books that are effectively entirely memorized (like *Harry Potter and the Sorcerer's Stone*), we do not expect this to be the case for LLAMA 3.1 8B, since it is a much smaller capacity model. Together, this will clarify how it is not the case that just making prefixes longer means that one could generate any target suffix they so desire—regardless of whether it was memorized. We also include LLAMA 2 13B. This model is a similar size as PHI 4, which we use in our negative controls. They have similar capacity, but were clearly trained very differently. While we do not know all of the differences, we do know that LLAMA 2 13B was trained on Books3, while PHI 4 was not. For one experiment on non-training data, we also include QWEN 2.5 72B as a point of comparison.

Procedure. We use the same sliding-window procedure as in Section 3.1, documented in detail in Appendix E.2. We inherit all of those setup details; the only difference is that we run for sequences of different lengths (and associated different prefix lengths) to extract

50-token suffixes: 75 tokens (25-token prefixes + 50-token suffixes); 100 tokens (50-token prefixes + 50-token suffixes, i.e., the typical setting); 150 tokens (100-token prefixes + 50-token suffixes); 250 tokens (200-token prefixes + 50-token suffixes); 450 tokens (400-token prefixes + 50-token suffixes); and, 850 tokens (800-token prefixes + 50-token suffixes). On our hardware, it takes approximately 13 hours to run *We Were Eight Years in Power* through LLAMA 3.1 70B with 850-token sequences (compared to $\sim$1.5 hours with our typical setup with 100 tokens, see Appendix E.3.33). This cost is the main reason we do not expand these experiments to a larger set of books and models.

F.3 Results of varying prefix length on known training data

We next provide extended results for baseline experiments that vary prefix lengths on known training data. Our results yield the outcomes we expect: we are able to surface more memorization with increased prefix length—to a point. Except in rare cases like *Harry Potter and the Sorcerer's Stone* and LLAMA 3.1 70B, in which the entire book is memorized, arbitrarily increasing prefix length does not make every training-data suffix extractable. This makes sense, as not all training data are memorized. These results also make clear that absence of extraction is not itself evidence that a sequence is *not* a member of the training data. Most sequences from `Books3` are not extractable even from models like LLAMA 3.1 70B, which memorize a lot. So the claim we make with respect to absence of extraction signal is in this one direction. When we make claims about training data membership, these are either based on ground-truth knowledge (e.g., knowing that `Books3` was in the training data and that 2025 books were not), or are based on comparisons to our negative controls (Appendix F.6).

We include results for 3 different models on 4 books from `Books3`, as documented in Appendix F.2. The largest model we test is LLAMA 3.1 70B (Figure 214). For a point of comparison with a much smaller capacity model, we also include the smaller model from the same family—LLAMA 3.1 8B (Figure 215) and LLAMA 2 13B (Figure 216), since this model is of a similar size as PHI 4 (Section 4, Appendix F.4.1).

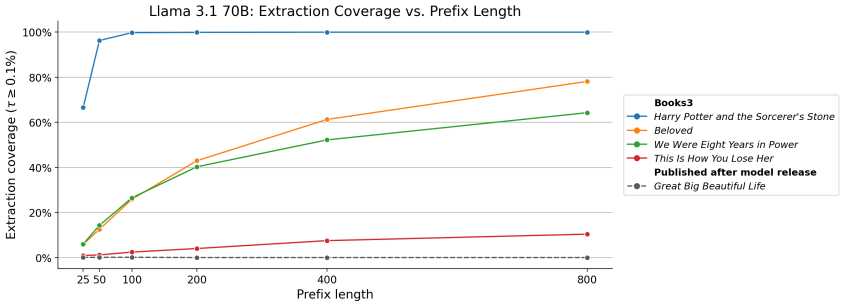


Figure 213: **Extraction coverage by prefix length.** Plotting extraction coverage for $\tau = \tau_{\text{min}} = 0.1\%$ (Equation 8) for the 5 books we test with varying prefix lengths and LLAMA 3.1 70B. Four of these books are in `Books3` and thus in LLAMA's training data. The remaining book, *Great Big Beautiful Life*, is used in our negative controls on non-training data; extraction coverage is 0% regardless of prefix length. As we increase prefix length, extraction coverage for each book in the training data begins to level off. While *Harry Potter and the Sorcerer's Stone* is entirely memorized by the model, it does not appear to be the case that *Beloved*, *We Were Eight Years in Power*, or *This Is How You Lose Her* would be entirely extractable with respect to 50-token suffixes and even longer prefixes.

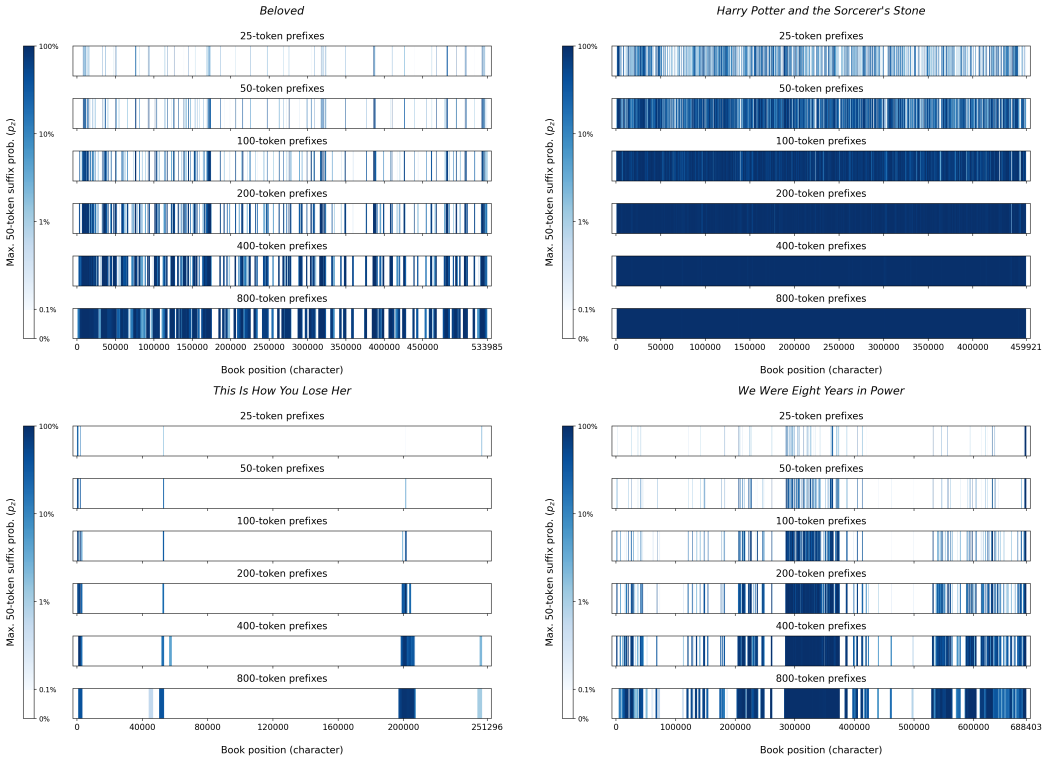


Figure 214: **Varying prefix for the LLAMA 3.1 70B baseline.** We run the sliding-window procedure for various prefix lengths for 4 books from `Books3`: *Beloved* (Morrison, 1987a), *Harry Potter and the Sorcerer's Stone* (Rowling, 1997), *This Is How You Lose Her* (Díaz, 2012), and *We Were Eight Years in Power* (Coates, 2017). LLAMA models are known to have been trained on `Books3`. We run probabilistic extraction on LLAMA 3.1 70B with top-k ($T = 1$, $k = 40$) decoding for 50-token suffixes and prefix lengths in $\{25, 50, 100, 200, 400, 800\}$. Extraction signal generally increases as prefix length increases. However, this only appears to be the case if there is indeed more memorization to surface; increasing prefix length does not always make previously unextractable sequences extractable. There are also diminishing increases as the prefix lengthens.

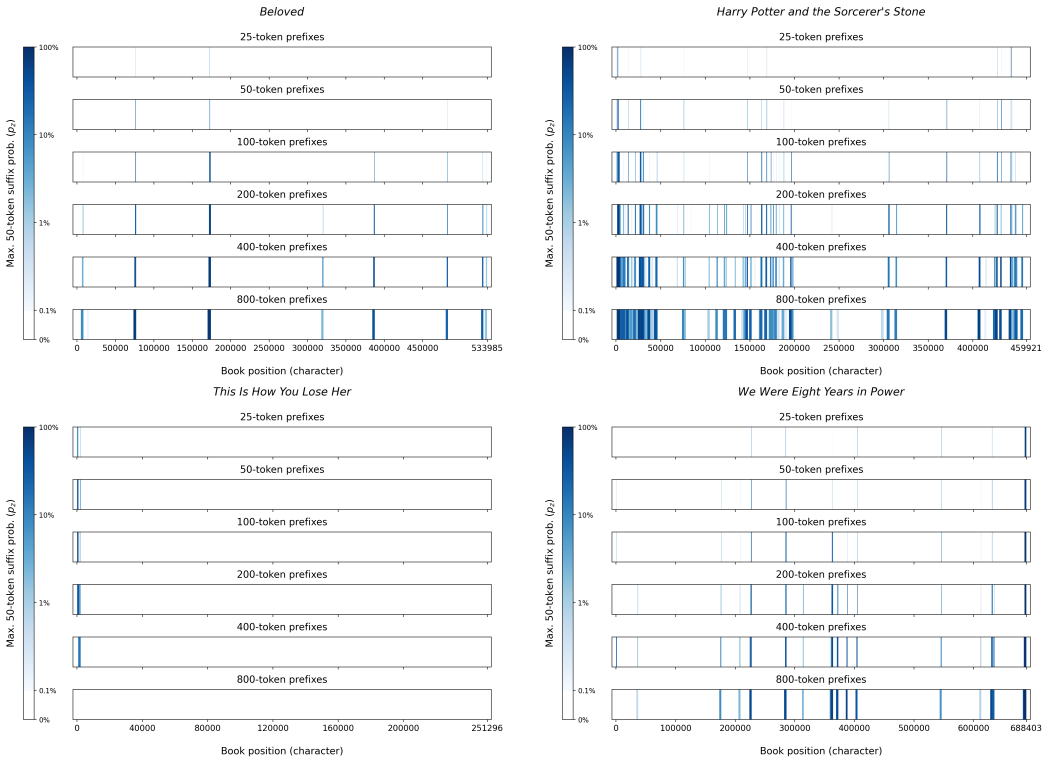


Figure 215: **Varying prefix for the LLAMA 3.1 8B baseline.** We run the sliding-window procedure for various prefix lengths for 4 books from `Books3`: *Beloved* (Morrison, 1987a), *Harry Potter and the Sorcerer's Stone* (Rowling, 1997), *This Is How You Lose Her* (Díaz, 2012), and *We Were Eight Years in Power* (Coates, 2017). LLAMA models are known to have been trained on `Books3`. We run probabilistic extraction on LLAMA 3.1 70B with top-k ($T = 1$, $k = 40$) decoding for 50-token suffixes and prefix lengths in $\{25, 50, 100, 200, 400, 800\}$. Extraction signal generally increases as prefix length increases. However, this only appears to be the case if there is indeed more memorization to surface; increasing prefix length does not always make previously unextractable sequences extractable. There are also diminishing increases as the prefix lengthens.

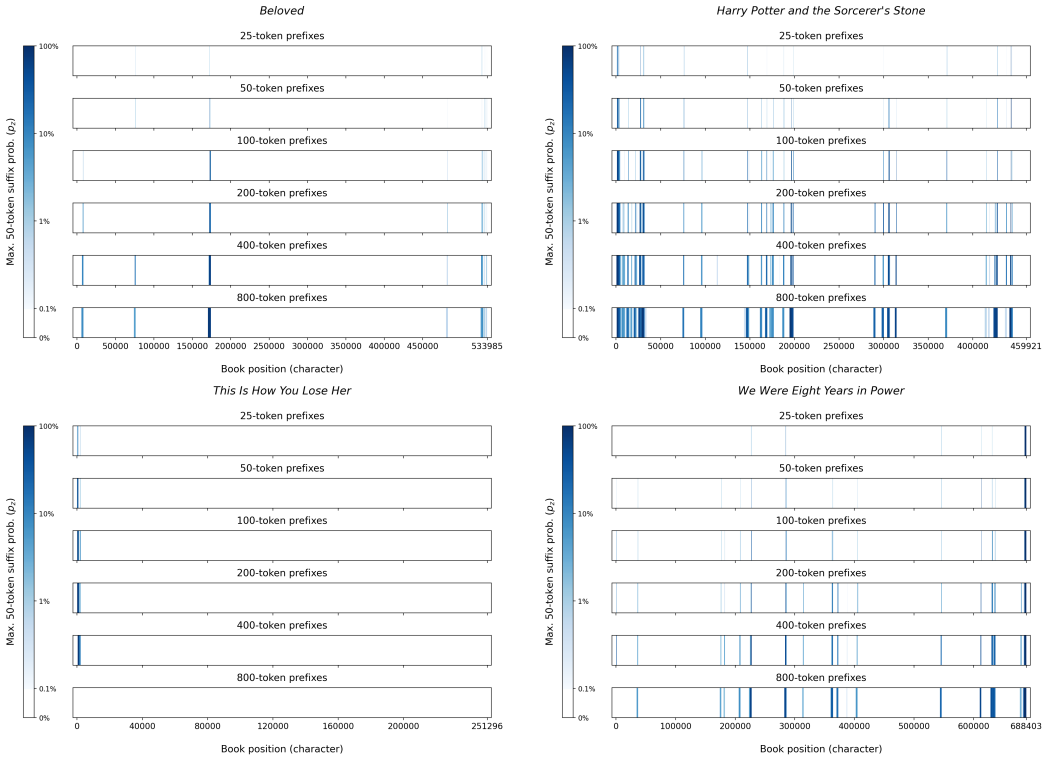


Figure 216: **Varying prefix for the LLAMA 2 13B baseline.** We run the sliding-window procedure for various prefix lengths for 4 books from `Books3`: *Beloved* (Morrison, 1987a), *Harry Potter and the Sorcerer's Stone* (Rowling, 1997), *This Is How You Lose Her* (Díaz, 2012), and *We Were Eight Years in Power* (Coates, 2017). LLAMA models are known to have been trained on `Books3`. We run probabilistic extraction on LLAMA 3.1 70B with top-k ($T = 1$, $k = 40$) decoding for 50-token suffixes and prefix lengths in $\{25, 50, 100, 200, 400, 800\}$. Extraction signal generally increases as prefix length increases. However, this only appears to be the case if there is indeed more memorization to surface; increasing prefix length does not always make previously unextractable sequences extractable. There are also diminishing increases as the prefix lengthens.

These results illustrate the high-level points described above. While for LLAMA 3.1 70B we are able to extract a lot more memorized sequences, we are able to extract relatively less from LLAMA 3.1 8B. That model has memorized less overall, which makes sense because it is a significantly smaller model. For LLAMA 3.1 70B and for some books, there is very little overall memorization that we can expose, regardless of prefix length. We can see this with the results in Figure 214 for *This Is How You Lose Her*. We also observe diminishing returns for increasing prompt length. This suggests that there is a maximum number of sequences we can extract from each work, even in the limit of setting the prefix length especially long. We can show this with plots of extraction coverage (Equation 8) as a function of prefix length, as in Figure 213 (duplicated from Figure 5 in the main paper).

F.4 Results of negative control experiments

We provide additional results for our varied-prefix-length negative control experiments (Section 4) on both PHI 4 (Appendix F.4.1) and non-training data that post-dates training cutoffs (Appendix F.4.2). While these experiments provide strong evidence for the validity of our procedure, we would want to run much more extensive experiments to responsibly lower τ_{min} or to make it LLM-specific. We defer this to future work, and here opt to pick a conservative global $\tau_{\text{min}} = 0.1\%$, which we discuss also in Section 4. The experiments that we run here are of a more substantial scale than the original probabilistic discoverable

extraction paper by Hayes et al. (2025b). That work performed one experiment on non-training data on a small PYTHIA model. The paper uses the results of this experiment as a basis to both validate the experiments on known training data in PYTHIA models, and to apply the methodology to much larger (and higher quality) LLAMA 1 and OPT models—i.e., to make claims about extraction of not-known-with-certainty training data, just as we do here with DEEPSEEK V1 67B, QWEN 2.5 72B, and GEMMA 2 27B.

Hayes et al. (2025b) finds that "…the number of queries n needed to generate unseen test data is orders of magnitude larger than for generating training data. [They] find that it is generally challenging to generate test data …This supports that, in [their] measurements of [probabilistic] discoverable extraction, matches between training-example targets and generated suffixes are almost surely due to memorization." We observe the same pattern, though instead of discussing it with respect to the number of queries n, as elsewhere, we frame our discussion in terms of p_z. In our results, the p_z that we observe on non-training data are many orders of magnitude lower than what we deem extraction success in our experiments on known training data. That is, our measurements for extraction are not erroneously capturing chance generation of non-training data. As in Hayes et al. (2025b), this gives us assurance that we can use our approach on models for which we do not know exactly what was included in the training data. We discuss this further in Appendices F.5 and F.6.

F.4.1 Results on PHI 4

We ran a large number of negative control experiments using PHI 4 (Abdin et al., 2024). As noted in Appendix E.1, this model was *not* trained on whole copyrighted books. It was trained on a mix of synthetic data, licensed websites, and public domain books. When running our sliding-window extraction procedure on this model for copyrighted books, we can expect (generally speaking, with important exceptions) that unique text from copyrighted books will not register as extractable with our measurements.

For our main results using the sliding-window procedure with 100-token sequences (Appendix E.3), we include PHI 4 as one of the 14 models that we evaluate. Here, in Figure 217, we show additional results for experiments on varied prefixes and the 4 in-copyright books from `Books3`. We are able to extract non-trivial amounts of public domain books from this model (e.g., Appendix E.3.61). In contrast, we are not able to extract much unique text from in-copyright books. As elsewhere, we extract boilerplate text like copyright notices and publishers' addresses. And we are also able to extract famous or otherwise popular quotes—either quoted in in-copyright books, or from the books themselves.

We (painstakingly) verify that every sequence that we extract from PHI 4 can be easily found in multiple sources on the Internet. That is, the quotes from books that we extract from PHI 4 are either in the public domain, quoted widely on blogs or platforms like Goodreads, or both. This is indicative (though not definitive) that these quotes could have found their way into PHI 4's training data from these online sources, rather from the books themselves.

For further support, we can also compare the text at the book locations for which we extract these sequences to heatmaps other models. We note that, almost always, these sequences are extractable (sometimes with enormous probability) from *all* models we test. Again, while this is not definitive, it is suggestive that this text is highly duplicated in sources that are a part of common pre-training corpora—that they are not necessarily memorized *because* they are in `Books3`, but because of their duplicated inclusion in the training data from other sources.

We also observe similar challenges—to a somewhat lesser extent—for our negative control experiments on books published after LLM training-date cutoffs. These books sometimes cite earlier sources—other books, news articles, social media posts, etc.—that pre-date the training-date cutoffs and are widely duplicated online. So, even this seemingly "clean" case also requires significant manual verification to confirm validity, which we discuss in Appendix F.4.2.

Altogether, it is reasonable to conclude that these are true instances of extraction. These results reinforce that our procedure detects true memorization when it exists and does not spuriously flag text where memorization is impossible. Even so, these results show

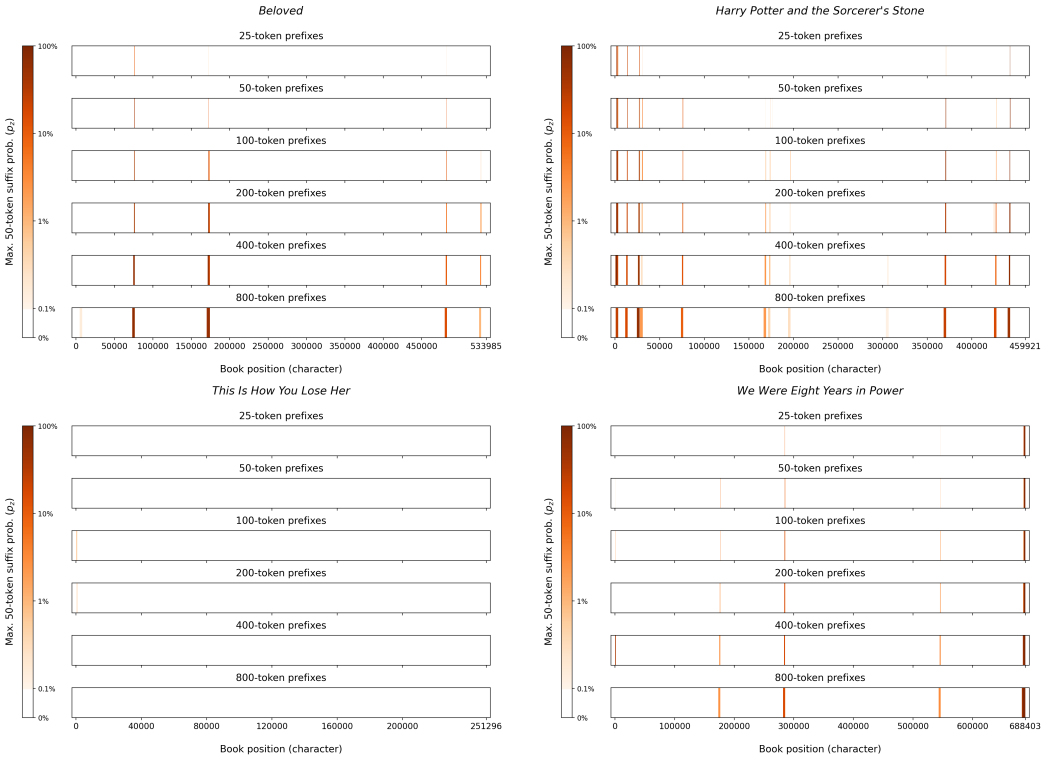


Figure 217: **PHI 4 negative controls.** Sliding-window procedure for various prefix lengths. We show results for 4 books from Books3: *Beloved* (Morrison, 1987a), *Harry Potter and the Sorcerer's Stone* (Rowling, 1997), *This Is How You Lose Her* (Díaz, 2012), and *We Were Eight Years in Power* (Coates, 2017). PHI 4 was deliberately not trained on whole copyrighted books. We run probabilistic extraction on LLAMA 3.1 70B with top-k ($T=1$, $k=40$) decoding for 50-token suffixes and prefix lengths in $\{25, 50, 100, 200, 400, 800\}$. Extraction signal generally increases as prefix length increases. However, this only appears to be the case if there is indeed more memorization to surface; increasing prefix length does not always make previously unextractable sequences extractable. There are also diminishing increases as the prefix lengthens. We manually verified that every extractable sequence comes from highly duplicated text that can easily be found on the Internet; such sources reasonably were a part of PHI 4's training data.

just how challenging it is to run completely clean negative controls for models trained on web-scraped data (Lee et al., 2023a), for which duplicates of the same piece of text across various sources complicate what training-data membership means. A given book may not have been a member of the training data, but this does not mean that all text from that book was not. In these cases, it seems extraordinarilyy likely that these sequences are members of the training data (and memorized). We discuss this further in Appendix F.6.

With the exception of the copyright notice, we are unable to extract any sequences from *This Is How You Lose Her* (Díaz, 2012). The sequences that we extract from *Beloved* are all posted (multiple times) on platforms like Goodreads. We discuss some of these quotes in Section 5 (Morrison, 1987b;c;d). *Harry Potter and the Sorcerer's Stone* (Rowling, 1997) is excerpted and quoted in an extraordinary number of places on the Internet. We were able to find every extracted sequence on 3 different web pages (after which we stopped searching/counting).

We discuss *When We Were Eight Years in Power* (Coates, 2017) in Section 4. As noted there, regardless of prefix length, there are only 4 areas—other than the copyright notice at the beginning of the book—that show extractable sequences. The sequences at the end of the book contain a list of ordered (page) numbers. This kind of list is a very common artifact from converting an EPUB file to TXT—as was done for the collection of Books3 (Reisner, 2023). Like the copyright notice, such ordered lists of numbers are text that, while high

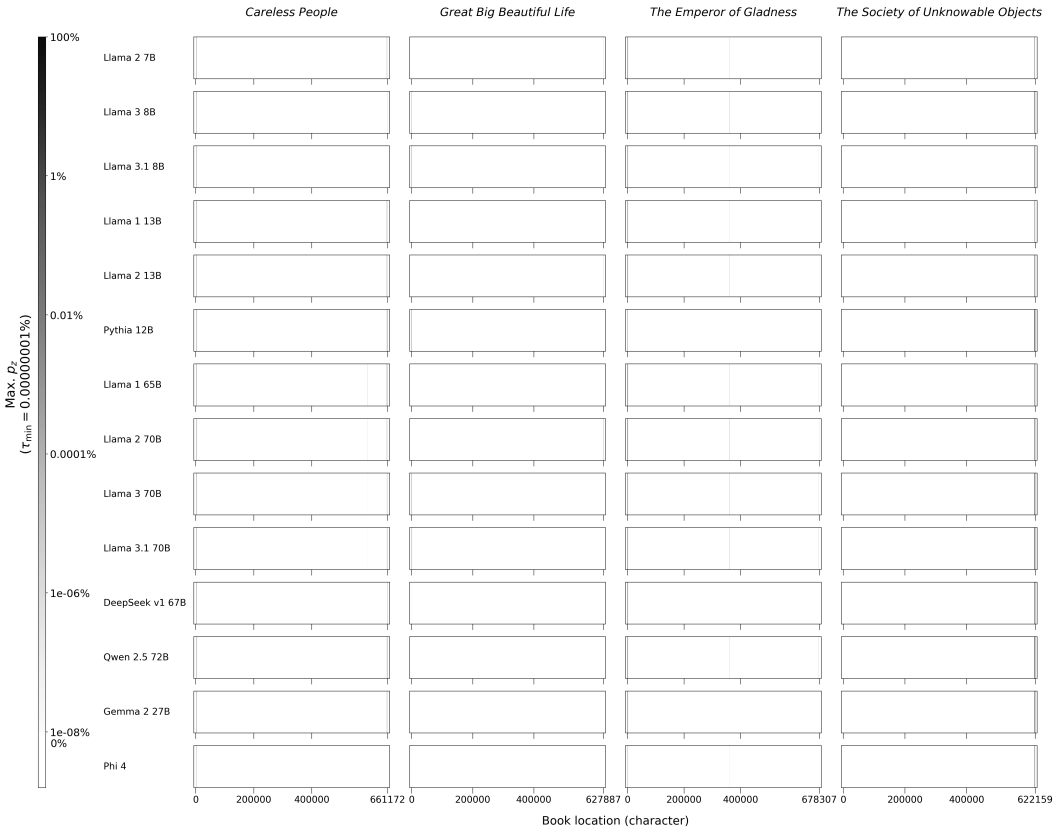


Figure 218: **Post-training cutoff negative controls.** We measure the probability of generating target 50-token suffixes for 50-token prefixes (using top-k, as elsewhere) for 4 books published in 2025: *Careless People* (Wynn-Williams, 2025), *Great Big Beautiful Life* (Henry, 2025), *The Emperor of Gladness* (Vuong, 2025), and *The Society of Unknowable Objects* (Brown, 2025). All of these books were published after the release dates for all models we evaluate. These experiments serve as a validity check that our measurements for probabilistic extraction are capturing true instances of extraction. By definition, no unique sequences in these books can be memorized because they are not members of the training data; we should not be able to (and do not) obtain high, non-zero probabilities—probabilities that we would consider to be successful instances of extraction for known or suspected training data. Note that, in contrast to our other heatmaps, we set a more permissive (i.e., lower) minimum for p_z ($p_z \geq 0.00000001\%$, or 1 in 10 billion generations), in order to see if there are sequences that can be generated even with fairly low probability that we would *not* count as successful extraction.

probability, we can discount as being particularly interesting with respect to our negative controls (Carlini et al., 2021). The other 3 are all text that is *not* unique to book, but rather highly duplicated text from the Internet (some of which is in the public domain):

- $\sim$176K: An exact quote from Malcom X's speech, "Who Taught You To Hate Yourself?" (May 5, 1962) (X, 1962)
- $\sim$280K: An exact quote of Deuteronomy 15:125 from the Bible – King James Version (KJV) (The Bible (King James Version)).
- $\sim$546K: A exact quote from Barack Obama's 2004 DNC Keynote: "hope of a young naval lieutenant bravely patrolling the Mekong Delta; the hope of a mill worker's son who dares to defy the odds; the hope of a skinny kid with a funny name who believes that America has a place for" (Obama, 2004)

All of these sequences are also highly memorized for other models we test.

F.4.2 Results on books from after training cutoffs

We run additional negative control experiments on 4 books that post-date the training of all models. As a result, these books could not have been included in the training data; we similarly should not expect to extract unique text from them for the different models we evaluate. However, as with our experiments on PHI 4, there are again cases where boilerplate text and popular quotes from other sources come up in our measurements.

Nevertheless, we do not detect extraction for *any* text unique to any of these 4 books for any model we test. In Figure 218, we show results for all 4 books using 100-token sequences (50-token prefixes + 50-token suffixes) for the 14 models that we include for our sliding-window experiments on `Books3` books (Appendix E). Note that in Figure 218, we expand the range of p_z that we plot to orders of magnitude below τ_{min}=0.1%. Even so, there are only a select few regions that register with non-zero probabilities that surpass τ_{min}, and they are duplicated generic text—in this case, the copyright notice—and some references drawn from pre-2025 sources. For example, for *The Emperor of Gladness* (Vuong, 2025), we extract a verbatim quote from the Bible (Mark 5:13). For *Great Big Beautiful Life* (Henry, 2025) and *The Society of Unknowable Objects* (Brown, 2025), we only extract boilerplate text.

For *Careless People* (Wynn-Williams, 2025), we extract the "careless people" quote from *The Great Gatsby* (Fitzgerald, 1925) (Figure 1). It is quoted at the beginning of the book (e.g., p_z=2.8% for LLAMA 3.1 70B):

> **Prefix**: to all this.\n&\nFor my grandmother Eileen\nWho regularly reminds us to "live an ordinary life" and "enjoy the good times."\nEpigraph\nThey were careless people, Tom and Daisy—they smashed up things and creatures and then
>
> **Suffix**: retreated back into their money or their vast carelessness, or whatever it was that kept them together, and let other people clean up the mess they had made.\n–F. SCOTT FITZGERALD, THE GREAT GATSBY\n

This is what we register as extracted close to the 0-character position in the heatmaps. (We highlight the portion from *The Great Gatsby* in blue.) The book also quotes Mark Zuckerberg in several locations—quotes that we confirm are often drawn from Facebook posts that are widely duplicated on social media platforms and in online newspaper articles. These make up the remaining extracted sequences. For example, one quote is from a 2016 post about a SpaceX rocket explosion (Wong, 2016) (e.g., p_z=13.5% for LLAMA 1 65B):

> **Prefix**: s SpaceX rocket explodes on the launchpad, completely destroying the Internet.org satellite it's supposed to be putting in orbit. Mark posted,\nAs I'm here in Africa, I'm deeply disappointed to
>
> **Suffix**: hear that SpaceX's launch failure destroyed our satellite that would have provided connectivity to so many entrepreneurs and everyone else across the continent. Fortunately, we have developed other technologies like Aquila that will connect people as well.

(We highlight the quoted portion from post in blue; even just a few words from the quote in the prefix are sufficient to generate the suffix with high probability.) Another is from a widely quoted (in news articles) Facebook press release (Lee, 2017):

> **Suffix**: people based on their emotional state.\nThe analysis done by an Australian researcher was intended to help marketers understand how people express themselves on Facebook. It was never used to target ads and was based on data that was anonymous and aggregated

An example of a sequence below τ_{min}—i.e., that we would not count as extractable, but that we include with our expanded range in Figure 218— for LLAMA 1 65 exhibits p_z=0.02% for a Mark Zuckerberg post about the Charlie Hebdo headquarters attack:

> need to reject—a group of extremists trying to silence the voices and opinions of everyone else around the world. I won't let that happen on Facebook. I'm committed to building a service where you can speak freely without fear of violence (Mlot, 2015)

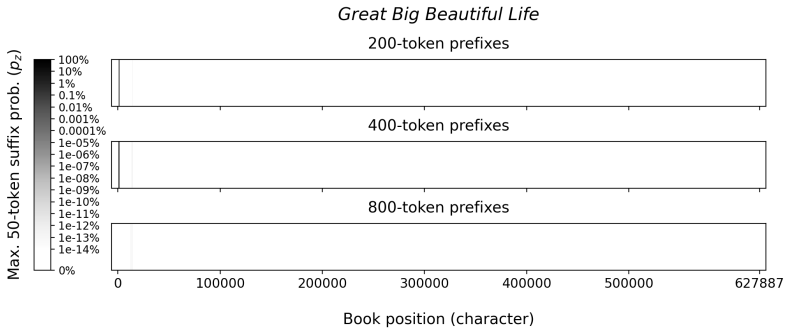


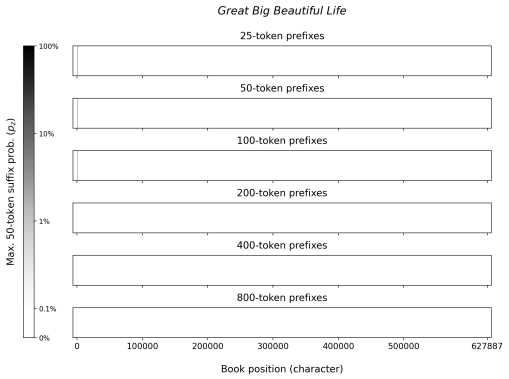
Figure 219: **Chance generation of non-training data.** We show LLAMA 3.1 70B and top-k ($T = 1$, $k = 40$) decoding on *Great Big Beautiful Life* Henry (2025) (published spring 2025), with respect to 50-token suffixes and prefix lengths in $\{200, 400, 800\}$ tokens. Except for the copyright notice at the beginning, we do not register any $p_z \geq \tau_{\text{min}} = 0.1\%$. We register no other suffix probabilities $p_z \geq 10^{-13}\%$, and only 1 (out of ∼62,000 sequences) for which $p_z \geq 10^{-14}\%$. (See the faint gray band near 20K characters.)

Mark Zuckerberg's manifesto is also quoted (Shahani, 2017).

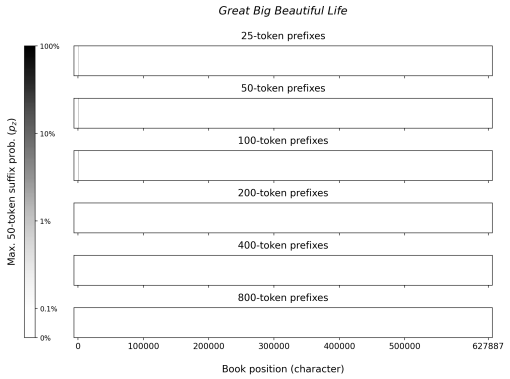
We know the *The Great Gatsby* quote is in `Books3`, and widely quoted elsewhere. Similar to our observations for PHI 4, it is reasonable to assume that these other highly duplicated quotes are similarly in the training data, and that these are true instances of extraction. Overall, consistent with Hayes et al. (2025b), we do not observe chance generation of non-training data until we significantly lower the probability threshold. In Figure 218, we observe no such instances for $p_z \geq 10^{-8}\%$, which is 7 orders of magnitude smaller than our chosen τ_{min}.

We also include the last book, *Great Big Beautiful Life* (Henry, 2025), in our varied prefix experiments as another point of comparison. We discuss these results for LLAMA 3.1 70B in Section 4 and in Figure 213, and provide the full set of results for 5 models in Figure 220. We extract no sequences (beyond boilerplate text) from any of the 5 models we test. In Figure 219, we show an extended p_z range to highlight how low we need to go beneath τ_{min} to register chance generation of non-training data. We extend to $10^{-14}\%$—13 orders of magnitude below τ_{min}=0.1%—and find exactly 1 sequence that surpasses this threshold for 200–800-token prefixes.

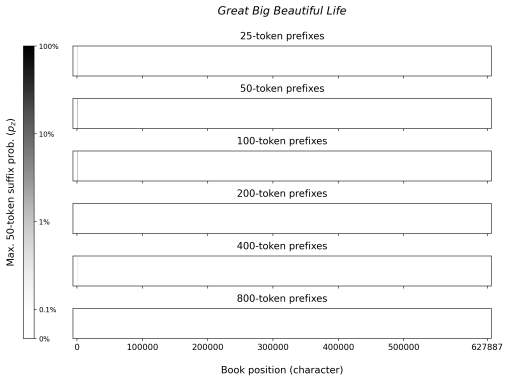
For example, for 800-token prefixes, we register 1 sequence with p_z=$6.06 \times 10^{-14}\%$. We report this as on the order of $10^{-14}\%$ in the main text because that is all we can reliably do for a number this small. The mantissa (6.06) is unreliable at this magnitude. This is because, even though we perform all log probability computations on CPU in full `float64` precision, producing the logits during inference on GPU (that we post-process on CPU) involves mixed-precision computations with the model weights/activations in `bfloat16` (Google Cloud Blog, 2019) and accumulation operations (matrix multiplications, normalization) in `float32` (Hugging Face, 2025). Each computation introduces small rounding errors that accumulate and propagate to the final logits. When we exponentiate to get probabilities, these errors result in only the exponent (not the mantissa) being reliable.



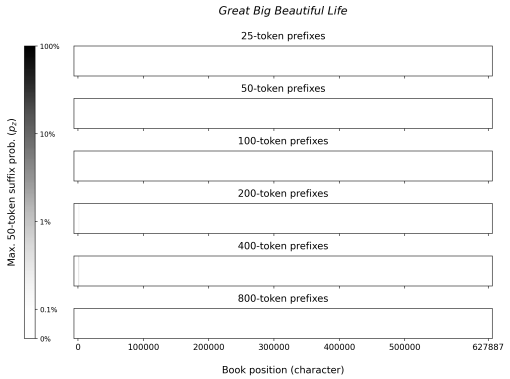
(a) LLAMA 3.1 70B



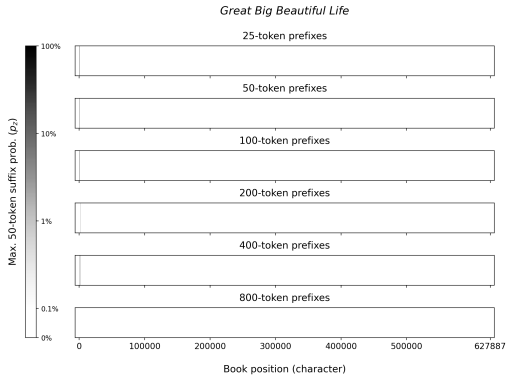
(b) LLAMA 3.1 70B



(c) LLAMA 2 13B



(d) PHI 4



(e) QWEN 2.5 72B

Figure 220: **Varied-prefix, post-training-cutoff negative controls.** We show results for *Great Big Beautiful Life* (Henry, 2025), which was published in spring 2025—after the release of all of the models we test. (The newest model is from the QWEN 2.5 series, which was released in late September 2024.) This book is not in any of the models' training data. We run probabilistic extraction on (**a**) LLAMA 3.1 70B, (**b**) LLAMA 3.1 8B, (**c**) LLAMA 2 13B, (**d**) PHI 4, and (**e**) QWEN 2.5 72B with top-k ($T=1$, $k=40$) decoding for 50-token suffixes and prefix lengths in $\{25, 50, 100, 200, 400, 800\}$. We are unable to extract any unique (e.g., non-copyright notice) sequences using any model or prefix length.

F.5 Limits on "the monkey at the typewriter" metaphor

A common trope is to compare an LLM to a "monkey at the typewriter" (Borel, 1913; The Simpsons: Last Exit to Springfield (Fox television broadcast Mar. 11, 1993)). The basic idea is that, given infinite attempts, an LLM could presumably output *any* continuation of a piece

of text. Like the monkey, it could by happenstance output a pristine copy of a Shakespeare play. This metaphor seems founded in the idea that LLMs are probabilistic. The monkey is randomly typing letters, and so, too, is an LLM, since there is an element of randomness in the generation process. For a given piece of text (the context), the LLM samples the next token to output based on the next-token conditional probabilities. That sampling is random, with respect to the conditional probability distribution of the possible next tokens.

Already here, this description of what an LLM is doing to generate text should start to sound very different from a "monkey at the typewriter." In that thought experiment, what the monkey is doing is random in the sense that *every* next character is equally likely to be produced at *every* step. But clearly, that is *not* what is happening with an LLM. The LLM has learned probabilistic relationships from the natural-language and other text (e.g., code) it was trained on. These relationships condition outputs given inputs. So while there is sampling involved—and thus probabilistic outcomes—those probabilities are not "random" in the same sense as the monkey typing. If an LLM could be used to produce a copy of a Shakespeare play (or *Harry Potter and the Sorcerer's Stone*, Section 6 & Appendix G), it would not be because it behaves randomly like the monkey. (One could set the temperature to an enormous number, effectively inducing "monkey" behavior via the decoding scheme, but this is not how LLMs are used in practice.)

A "monkey" or a "toothpick"? These probabilistic relationships can also be easily misunderstood in the opposite direction. One could imagine that a sufficient amount of context used as input would serve to bias the LLM's output to a particular, deterministic outcome: with a long enough prompt of the user's choosing, they could get an LLM to produce *any* particular, desired target output. This is the basic idea in Tyler Cowen's "toothpick argument" (Cowen, 2023):

> If you stare at just the exact right part of the toothpick, and measure the length from the tip, expressed in terms of the appropriate unit and converted into binary, and then translated into English, you can find any message you want. You just have to pinpoint your gaze very very exactly (I call this "a prompt").
>
> In fact, on your toothpick you can find the lead article from today's *New York Times*. With enough squinting, measuring, and translating.
>
> By producing the toothpick, they put the message there and thus they gave you NYT access, even though you are not a paid subscriber. You simply need to know how to stare (and translate), or in other words how to prompt.
>
> So let's sue the toothpick company!

The "toothpick" misunderstanding of an LLM is mutually exclusive with the "monkey" misunderstanding. For the toothpick, prior context has a determinative impact on the output; for the monkey, there is no context long enough that could ever have a determinative impact on the output. Note that in the toothpick metaphor, it should also be possible to get an LLM to generate outputs that are exact copies of *non*-training data—in this case, "the lead article from *today's* [paper]" (emphasis added). (One could, for an open-weight model, manually bias logits. This could force an LLM to produce certain outputs, effectively inducing "toothpick" behavior, but this is also not how LLMs are used in practice.)

Where our results come in. There is plenty to say about misplaced metaphors about generative AI. We defer to others who have written extensively about this (Lee et al., 2023b; Cooper et al., 2023; Cooper & Grimmelmann, 2024). Here, we highlight how our specific results make specific points clear about why these two particular metaphors are both misguided. The key point is that our experiments show that (a) there will be at least some target suffixes that could *never* be generated by an LLM, even with infinite prompts to the model with a given prefix, and (b) a specific desired target suffix cannot *always* be produced by an LLM, given a sufficiently long prompt. The evidence for both comes from the same results: for a given LLM θ and decoding scheme ϕ, some suffixes are *impossible* to generate, regardless of (a) how many times one independently prompts the model with a given prefix or (b) how long of a prefix one uses to condition the output.

Our results make this very clear in practice, but we can also make this point conceptually. Intuitively, for a sequence z, if the LLM θ with decoding scheme ϕ assigns the verbatim suffix (given the prefix) zero probability (i.e., $p_z{=}0$), then no amount of independent sampling trials will ever produce that suffix. If instead $p_z > 0$, then repeated independent sampling ensures that the probability of seeing z at least once tends to 1 as the number of trials grows. Therefore, for any finite set of sequences $\mathbb{Z}$, the maximum number of verbatim suffixes that can ever be generated is exactly those with nonzero probability:

$$\lim_{n\to\infty} \mathbb{E}[\text{\# unique verbatim suffixes produced}] \;=\; \big|\{z \in \mathbb{Z} : p_z > 0\}\big|.$$

Proof. For a single sequence z with per-trial, verbatim-suffix generation probability p_z, the probability of generating the suffix at least once after n independent trials is

$$\Pr(\text{verbatim suffix produced in } n \text{ independent trials}) \;=\; 1 - (1 - p_z)^n.$$

If $p_z = 0$, this expression is 0 for all n. If $p_z > 0$, then $(1 - p_z)^n \to 0$ as $n \to \infty$, so $\Pr(\text{verbatim suffix produced in } n \text{ independent trials}) \to 1$. Now let $\mathbb{Z}$ be a finite set of sequences. By linearity of expectation,

$$\mathbb{E}[\#\text{verbatim suffixes produced in } n \text{ independent trials}] \;=\; \sum_{z\in\mathbb{Z}} \big[1 - (1 - p_z)^n\big].$$

Taking $n \to \infty$,

$$\lim_{n\to\infty} \mathbb{E}[\#\text{verbatim suffix produced in } n \text{ independent trials}] \;=\; \sum_{z\in\mathbb{Z}} \mathbf{1}[p_z > 0].$$

□

Therefore, for this fixed finite set $\mathbb{Z}$, there is a maximum attainable suffix (given prefix) generation rate: the fraction of $\mathbb{Z}$ whose suffixes can ever be generated. This limit is independent of the target success level p in (n, p)-discoverable extraction (Definition A.1). As n grows large, the rate reflects not only extractable (memorized) sequences—i.e., those with sufficiently high probability such that we consider them to be instances of extraction (in this work, $\tau_{\text{min}} \geq 0.1\%$)—but also any z with $0 < p_z < \tau_{\text{min}}$ that we would not count as extraction.

Sequences that cannot be generated in practice. There are various ways that a given suffix could have $p_z = 0$. It might be the case that, for a particular decoding scheme, it is impossible to sample certain tokens, making some suffixes unreachable. This is the case for top-k decoding, which only considers the top-k-highest probability tokens at each sampling iteration. Any token not included in the top-k tokens forecloses possible output generation paths (Appendix A.3). For $k = 40$ and for LLAMA 1 models, only 40 out of 32,000 tokens could be generated in a given sampling iteration.

For another, numbers are represented in computers with a finite number of bits. This means that there is a minimum number that can be represented (> 0) before it gets rounded down to 0. For instance, imagine a token vocabulary of 32,000 tokens (as with LLAMA 1 models). If the model just generated tokens completing randomly (by setting temperature very high, like a "monkey at the typewriter"), then at each iteration every token would have probability 1/32,000. When generating 71 such tokens in a row (i.e., a 71-token suffix), the probability of any such suffix is $(1/32{,}000)^{71} \approx 1.3626 \times 10^{-320}$. Once we try to generate a 72nd token, the probability for the whole sequence becomes $(1/32{,}000)^{72} = 0$: in terms of how the probability gets represented in the computer, there is underflow (i.e., rounding down to 0). As a result, even with completely-random temperature sampling, not every sequence could possibly be generated.

We often observe 0-probability suffixes regardless of prefix length. This is true for non-training data (Section 4, Appendix F.4.2) and for training data that are not memorized (Section 4, Appendix F.3). Even if we add more context to try to condition it in a particular direction—there are many cases where we simply are unable to raise p_z above 0 (let alone above τ_{min}, the minimum probability we consider evidence for extraction).

F.6 Additional notes on extraction and membership inference

There are several points that are worth addressing with respect to the broader literature on **membership inference** and **extraction**. Here, we briefly situate our work—particularly our work with respect to validity—in relation to prior work. This is important to do for clarity reasons, because our work and the types of claims we make are very different from prior work. Prior work tends to focus on *average* extraction rates over *random* samples of known or proxy training data. Since we are making claims about *specific* works from a *specific* population of text, the ways that we go about evaluating and validating our work are necessarily different. This is not just a detail or aside; it has to do with the entire purpose of this project, and significantly influences our measurement approach in ways that contrast with other work on membership inference.

F.6.1 The relationship between membership inference and extraction

As noted in Carlini et al. (2021), "Given a set of samples from the model, the problem of training data extraction reduces to one of membership inference: predict whether each sample was present in the training data." This makes sense intuitively. Extraction is defined with respect to memorization: it is only possible to extract memorized training data, and (tautologically) training data can only be memorized if they are included—i.e., are **members**—of the training dataset. To demonstrate extraction is therefore to demonstrate memorization, and memorization implies membership.

Not only does extraction imply membership of a piece of data, it actually reproduces that data in model outputs. In this respect, it is a stronger evidentiary claim (and, at least in principle, a harder one to make): membership inference asks a one-bit question—was this sequence a member of the training data?—while extraction produces the sequence itself. (At the same time, reliable membership inference with even the strongest **membership inference attacks** is extraordinarily challenging for LLMs (Hayes et al., 2025a), while extraction is relatively straightforward.) Put differently, one can make membership inference claims *without* extraction (e.g., with classifiers, see below); these claims can involve evidence where extraction might fail. So, even though successful extraction always implies membership, this does not go in both directions: membership does not imply successful extraction (Appendix F.6.5).

Carlini et al. (2021) goes on to note that "In their most basic form, past membership inference attacks rely on the observation that models tend to assign higher confidence to examples that are present in the training data. Therefore, a potentially high-precision membership inference classifier is to simply choose examples that are assigned the highest likelihood by the model." At a high level, this is clearly directly related to extraction. Greedy-decoded discoverable extraction approximates the highest-likelihood sequence (albeit pretty poorly) with greedy decoding (Carlini et al., 2023). Probabilistic extraction does the same but allows more flexibility than greedy decoding, and can identify sequences with higher likelihood than greedy-decoded ones (Hayes et al., 2025b).

F.6.2 We know ground-truth membership for most of our experiments

In much work on extraction, researchers do not demonstrate extraction to *infer* membership. Instead, they run extraction experiments on *known* training data (Hayes et al., 2025b; Carlini et al., 2023; Lee et al., 2022; Hayes et al., 2025a), i.e., where ground-truth membership is not in doubt. This makes sense as a setup for drawing valid conclusions about extraction and memorization: membership is a prerequisite for memorization, and memorization is a prerequisite for extraction. Testing extraction on known member sequences therefore enables straightforward conclusions about memorization. Here, memorization is the most plausible explanation for extraction. As Carlini explains, when a sufficiently large and unique piece of training data is extracted, "the only possible explanation is that the model has somewhere internally stored [that piece of training data]" (Carlini, 2025). As Cooper and Grimmelmann similarly note, "in order to be able to extract memorized content from a model at generation time, that memorized content must be encoded in the model's parameters. There is nowhere else it could be. A model is not a magical portal that pulls fresh information from some parallel universe into our own" (Cooper & Grimmelmann, 2024, p. 25). (See Section 2.)

Most of our experiments fall into this category, as the large majority of our results are for LLAMA and PYTHIA models, where it is known that Books3 was included in the training data (Sections 3–5; Appendices E.1 & F.2). This is precisely the motivation behind our baseline experiments (Section 4, Appendix F.3), which we use as a basis of comparison for our negative controls (Section 4, Appendix F.4). Not only do we know that these models were trained on the books in Books3, we know that they were trained precisely on the Books3 corpus.

Nevertheless, while most work on greedy-decoded discoverable extraction does not do this, we perform validity experiments anyway. We need to set a minimum valid extraction probability τ_{min} (Appendix F.1). In this paper, we use a very high global τ_{min} to avoid false positives, and we validate this choice with our negative controls. Hayes et al. (2025b) do not set an explicit τ_{min}, and do not make strong claims about when probabilistic extraction is valid; however, they do run a negative control experiment to show more generally that probabilistic extraction is a reasonable metric.

In the original version of this paper, we included the negative controls on PHI 4 (Appendix F.4.1) to support our claims. Combined with the extensive analysis we perform on known training data for LLAMA, this is how we chose a global τ_{min}: a conservative threshold that would not over-count extraction (and thus memorization). Even before including our experiments on post-cutoff books, our results provided ample evidence that non-training data exhibit 0 or orders of magnitude smaller probabilities p_z than training data. The post-cutoff negative controls provide further confirmation.

This was true both for smaller, lower quality models, and higher quality models that exhibit lower loss (on average). While larger models like LLAMA 3.1 70B fit their training data better than smaller ones, and exhibit lower loss on average, they do not (and cannot) exhibit low loss everywhere, such that our measurements of log probabilities would be confounded. This is also clear from our experiments on known training data—Books3 books. For many books and large models, we are unable to extract any unique sequences; there are effectively no nonzero sequence probabilities. It is plausible that future work will find lower, model-specific τ_{min} thresholds, which may perhaps need to be set higher for larger models than for smaller ones. For now, however, the threshold we set is so conservative that these nuances do not affect our conclusions.

F.6.3 Inferring membership for LLMs with undisclosed training data

Given our extensive results on known training data—including experiments to vet their validity—we can apply our methodology with confidence to models whose training data are not known with certainty. We do this for three model families: DEEPSEEK V1, QWEN 2.5, and GEMMA 2. For these models, validated extraction serves as evidence of membership. These experiments differ from those on models known to be trained on Books3, but we ensure their validity through extensive negative controls and baselines.

This approach is consistent with prior work. Carlini et al. (2021) performed experiments on GPT-2. Nasr et al. (2023; 2025) extracted training data from open-weight models with undisclosed training data and from ChatGPT, a closed system. In both cases, the authors relied on proxy datasets and strategies to reduce bias in their results, such as shorter prompts. More generally, this is an important and reasonable thing to try to achieve. As researchers, we should be developing valid techniques for extraction that are useful in cases for which we do not know ground-truth membership. If membership inference required knowing membership in advance, there would be nothing to infer; at best—when it works—it would confirm what we already know.

We do not need to use shorter prompts in our work because we validate our work with extensive negative controls. Further, we manually verified that every sequence extracted from PHI 4 and from post-cutoff books was widely duplicated online. Our validity experiments go beyond most prior work, which often does not include negative controls. This is intentional; our work has the potential to impact copyright litigation, not just the scientific literature. We therefore believe it is important to hold our work to a higher bar. Importantly, since training data for these models are undisclosed, we make claims only at the sequence level, not at the level of whole books (Appendix F.6.4).

Finally, the population of data we study differs from most prior work, which evaluates memorization broadly across diverse training data (websites, code, spam, etc.) (Carlini et al., 2021; 2023; Nasr et al., 2023; 2025; Hayes et al., 2025b). Such work must account for false positives like "trivial memorization" (e.g., number lists) or "repeated substrings" (e.g., "I love you" loops). LLMs assign high probability to both of these types of sequences, even if they are not memorized. We note these phenomena explicitly where we identify them, but they are generally not relevant our domain of long-form narrative text.

The rare cases we identify of trivial memorization—lists of page numbers (an artifact of converting EPUB to TXT), chapter indices—are excluded from our claims about extraction of unique creative expression. We identify one case of repeated substrings. Silverman (2010) exhibits elevated probability in a part of the book that repeats "I will not wet the bed." four times in a row. But this sequence does not surpass τ_{min}, anyway, so we do not count it as extracted. (We also exclude boilerplate instances of true memorization from our claims, like copyright notices and author biographies.) Across millions of sequences, these account for only a few hundred cases. This highlights why the population under study matters for validity: our focus on books yields different considerations than prior work, and we address these systematically. Unlike the work done in Carlini et al. (2021), these types of edge cases are extremely rare in our population of interest.

Prior work on copyrighted text and LLMs. In the technical literature, there are a few earlier papers that address the generation of copyrighted text by LLMs. While this work (e.g., Karamolegkou et al. (2023)) is important initial work, for the reasons discussed in this appendix, much of it arguably does not directly study memorization, even if it is framed in those terms. In particular, without ground-truth information about training data membership and without negative controls to calibrate false positives, it is difficult to interpret reported overlap-based results as valid evidence of extraction (and thus memorization). (Further, in the case of Karamolegkou et al. (2023), results for very short-form generations—which are insufficient on their own to support extraction claims—are presented alongside longer recovered text within the same plots, making it difficult to distinguish stronger from weaker forms of evidence.)

F.6.4 We infer membership for sequences, not whole books

We emphasize that our claims about membership are made at the level of suffixes, not entire books. For models such as DEEPSEEK V1, QWEN 2.5, and GEMMA 2, where training data are undisclosed, we cannot establish whole-book membership. Instead, our extraction results are about specific sequences drawn from `Books3`, a known pretraining dataset, and it is at this level that we make membership claims. In some cases, our results are suggestive that entire books were included in training, given that we can extract high-probability suffixes spanning a full work (e.g., Appendix E.3.153), but we do not generalize from suffixes to whole-book membership.

We also do not claim that we know the provenance of those extracted sequences. For LLAMA models we know that the training data included `Books3`, and thus at least one copy of each sequence from the books we test was included in the training data via `Books3`. But this is about all we do know with certainty. We do not know very much about these models' training data details more generally; knowledge of the inclusion of `Books3` is a rare exception, not the norm. Books sequences may have been included in the training data via other unknown sources: other book corpora, widely quoted excerpts in web scrapes, etc. The same could be said for these other models, where we do not know if `Books3` was included in the training data.

Altogether, these observations highlight that membership questions for LLMs are rarely straightforward. This is already widely known in the literature, with respect to the presence of duplicates in training datasets (Lee et al., 2022; Carlini et al., 2023; Hayes et al., 2025b). This discussion illustrates how these challenges manifest specifically for books, and why membership claims should be scoped carefully. Our measurements are at the level of individual suffixes, but that is not necessarily the only level at which membership is of

interest. For copyright, membership may be meaningful in different ways at different scopes—an entire book, a chapter, a single sentence, etc. (Lee et al., 2023b).

F.6.5 Absence of extraction does not imply anything about membership

We state this with respect to describing the problem of membership inference in Appendix F.6.1, but it bears repeating: extraction implies membership, but membership does not imply extraction. This is clear in our results; many books in Books3 have no extractable unique text for LLAMA models, even though those models were trained on Books3. So when we see $p_z = 0\%$, we should not conclude that a given sequence (or the source that contains it) was not included in the training data. This is trivially true for LLAMA models, but more generally true for models with undisclosed training data. For example, QWEN 2.5 may have been trained on Books3; just because it exhibits low extraction probabilities on Books3 books is not evidence that it was not trained on this corpus. Further, other extraction techniques than the ones we use here could possibly surface additional memorization; so absence of extraction *with our measurement procedure* is not definitive that a particular sequence was not memorized.

Further still, $p_z = 0\%$ does not necessarily imply anything about how much a particular sequence may have influenced the model or its outputs. Attribution of generations to training data is a different problem, both with respect to technical work in ML and implications for copyright. We do not study these topics in this work. At a very high level, one can see this distinction through our brief results in Appendix G.3. We run our reconstruction procedure on *Sandman Slim* (Kadrey, 2009) and LLAMA 3.1 70B. This model does not memorize this book at all, with respect to our extraction procedure. We run this experiment to show that it fails—that reconstruction is not something that trivially works for any book. Nevertheless, even though we fail to reconstruct any text from the book, it is clear that the generation we produce was influenced by the text (in an informal sense). This type of influence is not something that we capture or attempt to capture—either formally or informally—in our measurements in this paper.

F.6.6 Extractability and copyright

Separate from the above technical considerations about membership and extraction, there are other considerations for copyright. A sequence can count as memorized in a technical sense—in this paper, when $p_z \geq \tau_{\text{min}} = 0.1\%$—but this does not necessarily mean that all types of memorization are relevant to copyright. As we discuss in Section 7 and Appendix B, sequences that have very low extraction probabilities may matter less for copyright than those with high ones, since the latter are far easier to extract reliably in model outputs.

G Recovering books with one seed prompt

Given the extent of memorization that we observe for *Harry Potter and the Sorcerer's Stone* (Rowling, 1997) (as well as other books) and LLAMA 3.1 70B, it seemed possible that we could recover the entire book near-verbatim using only a single seed prompt of ground-truth text drawn from the book.

This result was a natural follow-on to our memorization experiments, but it is not the main focus of our paper. Therefore, this part of our paper reflects only a preliminary investigation into recovering large book segments from an LLM. It is an existence proof that this type of result is possible, not a comprehensive investigation into how to make this type of method work for different texts.

G.1 Recovery method

We successfully recover the book near-verbatim with a variety of seed prompts and variations on the same strategy. Overall, we observe that this result is stable—across different hardware, and for different settings. However, the procedure is relatively sensitive to how it is configured; for example, if too long of a stride is used, then it we observe divergence

from the true text. We have preliminary results that significantly improve the stability of this approach. We defer investigating this to future work.

Seed prompt. We obtained high-fidelity recoveries of the book with several seed prompts, including the first 390 tokens of the the first chapter (the first few paragraphs) and the first line of the first chapter (60 tokens):

> Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much. They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with such nonsense.

With some additional exploration of how to configure generation, we also were successful with prompting with just "Mr. and Mrs. D" and "Chapter ONE:\nTHE BOY" (both 7 tokens). The former is sufficient for the model to output "ursley" and then the rest of the first line; the same is true for the latter, which outputs the rest of the first chapter's title ("THE BOY WHO LIVED"), and then proceeds to output the first substantive line of the book. (These token counts include the first token, which is the beginning-of-sequence—`BOS`—token.)

Overall approach. We take a sliding-context approach that we discuss in Section 6. We take an initial seed s and top it up to a minimum of 50 tokens (if it is less than that). On some hardware, for stability, we normalize the format of that longer seed (e.g., making sure that em dashes are consistently represented) before proceeding to long-form generation. In all cases, we do not guide the procedure beyond providing the initial context. We use beam search to generate 50 tokens at a time, and use a maximum context window of 3000 tokens for each generation. This means that we append the 50 tokens we generate to the existing context, and clip the first 50 tokens for iterations that exceed the maximum context. (This is most of the iterations; only at the start do we have a context that can be less than 3000 tokens).

We succeeded with 9 tries in total. (Our first major success occurred in the third run, when we recovered ten chapters—chapters 1-9, and 12—near-verbatim.) The first 3 runs were debugging and setup (e.g., testing the key idea with greedy decoding and getting a sense of how short a reasonable seed prompt could be). The last 6 involved changing the size of the number of maximum context tokens (we settled on 3000) and the number of beam candidates (8), as well as minor tweaks to handling `EOS` tokens.

Handling `EOS` tokens was the only (minor) complication, and it was simple to resolve. `EOS` tokens tend to be predicted at the ends of chapters; we therefore remove them and manually replace them with `"CHAPTER {n+1}"`, with `{n+1}` spelled out (e.g., `"TWO"`, `"THREE"`). Other attempts involved `"Chapter"` (instead of all caps) and using the number `{n+1}` (e.g., `"2"`) instead of spelling out the chapter number (e.g., `"TWO"`). Other solutions could just down-weight the logits for `EOS`.

Occasionally, the model does not predict an `EOS` at the end of a chapter, which leads to some misalignment later on (e.g., inserting `"CHAPTER FOUR"` for the fifth chapter). This can cause the model to repeat segments of chapters that have already been produced. The fix here was also simple. We keep track of how many tokens have been generated since the last `EOS` token, and if that number surpasses 10,000, we assume that we have missed the end of a chapter and account for this (i.e., move the chapter counter ahead even though we did not see an `EOS`). In future runs, we were able to remove this logic entirely. We defer discussion to future work.

The exact code that we ran using the first 390 tokens as the seed prompt is in Listing 1. Since this code involves beam search, our results should be (in theory) deterministic. Because in practice hardware non-determinism can complicate this, we ran a (non-exhaustive) test by executing the code on two different sets of 4 GPUs. Both yielded the same exact output. (Though we also spot-checked some of the underlying logits, and there were some slight differences.)

Comparing to the ground-truth text. The substantive diff between the version of *Harry Potter and the Sorcerer's Stone* (Rowling, 1997) that we have from `Books3` and our generated

output is minimal. We do some very minor text normalization of the ground-truth book. That is, we trim the beginning and end of the ground-truth book to not contain front matter or anything after the end of the last chapter. We remove _ from both documents (used for italics in the Books3 version) and align ellipses to "..." (which appear as " . . ." in the Books3 version.) We then compute various similarity metrics betwen the two documents. The cosine similarity of TF-IDF vectors of each document is an astounding 0.9999. TF-IDF is a limited metric, as it treats documents as bags of words and thus fails to capture word order. So, as two additional points of comparison, we compute similarity using greedy longest common substring (LCS) matching at both the word and sentence levels (via `difflib`'s SequenceMatcher). A score of 1 indicates a perfect match and 0 indicates no match. We obtain a word-level similarity of 0.992 and a sentence-level similarity (which is more sensitive to formatting differences) of 0.934.

The qualitative differences involve minor formatting changes (that we do not normalize) and localization differences. For example, the Books3 version has British spelling (e.g., "Mummy" instead of "Mommy"). These formatting and localization differences are the large majority of the diff between the two texts. Occasionally, the model also misses single lines, e.g., "There was a loud 'Oooooh!"' (Figure 7).

One of the most interesting outcomes (that we have not investigated in detail) is very occasional apparent hallucinations. For example, in the first chapter, we observe the following diff:

> "Yeah," said Hagrid in a very muffled voice, ~~"I'd best get this~~ **"I'll be takin' Sirius his** bike ~~away.~~ **back.** G'night, Professor McGonagall — Professor Dumbledore, sir."

Figure 221: **Portion of the diff from the first chapter of *Harry Potter and the Sorcerer's Stone*.** Crossed-out red text is in the ground-truth book in Books3, but was not generated by our recovery procedure; text highlighted in yellow/green is in the recovered text but not the ground-truth text.

Selecting the top sequence candidate with beam search always led to this particular outcome in Figure 221, regardless of how we configured our method. Given how specific this particular change was, we investigated it further. It turns out that the generated text reflect the ground-truth from editions of the book printed before 2004. In later editions, this line was changed to the one that we do not generate—the one that is in the ground-truth version we have access to in Books3. This change was done for plot continuity reasons, based on an inconsistency that emerged due to plot development concerning Sirius Black in the later books (Harry Potter Wiki). This line is not actually hallucinated; it is just from an earlier edition of the book than the one we diff with. Of course, this difference still contributes to our quantitative measurements of similarity, even though it is not a meaningful substantive difference.

```
from transformers import AutoTokenizer, AutoModelForCausalLM
import torch, json
from tqdm import tqdm

# load the first line of HP 1
with open("seed.txt", "r", encoding="utf-8") as f:
    seed_text = f.read().strip()

# Load the model and tokenizer
model_name = "meta-llama/Llama-3.1-70B"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForCausalLM.from_pretrained(
    model_name,
    device_map="auto",
    torch_dtype=torch.bfloat16,
)
model.eval()

max_context_tokens = 3000
max_new_tokens = 50
# the book is more like 106k;
# adding buffer for potential formatting differences
max_story_tokens = 113000

```

```
eos_token_id = tokenizer.eos_token_id

input_ids = tokenizer(seed_text, return_tensors="pt").input_ids.to(model.device)
generated_ids = input_ids

# Set up progress bar
pbar = tqdm(total=max_story_tokens, desc="Generating story tokens")
pbar.update(generated_ids.shape[1])

generation_steps = []
generation_num = 1
chapter_count = 1

chapter_nums = {1 : "One", 2 : "Two", 3 : "Three", 4 : "Four", 5 : "Five", 6 : "Six", 7 : "Seven", 8 : "Eight", 9 : "Nine", 10 : "Ten", 11 : "Eleven", 12 : "Twelve", 13 : "Thirteen", 14 : "Fourteen", 15 : "Fifteen", 16 : "Sixteen", 17 : "Seventeen"}

tokens_since_last_eos = 0

while True:
    # Get the tokens for the (up to) 3000-token prompt
    slice_start = max(0, generated_ids.shape[1] - (max_context_tokens - max_new_tokens))
    input_ids_window = generated_ids[:, slice_start:]
    attention_mask = torch.ones_like(input_ids_window)

    # call generation code
    outputs = model.generate(
        input_ids_window,
        attention_mask=attention_mask,
        max_new_tokens=max_new_tokens,
        do_sample=False,
        num_beams=8,
        early_stopping=False,
        length_penalty=1.2,
        temperature=1.0,
        top_p=1.0,
        pad_token_id=eos_token_id,
    )

    # remove the prompt from the output
    new_tokens = outputs[:, input_ids_window.shape[1]:]
    new_tokens_list = new_tokens[0].tolist()

    tokens_since_last_eos += len(new_tokens_list)

    # Some fun tricks to deal with EOS tokens
    if eos_token_id in new_tokens_list:
        if tokens_since_last_eos >= 10000:
            print("More than 10000 tokens since last EOS; probably missed a chapter break")
            print("Incrementing chapter count")
            chapter_count += 1

        print(f"EOS: replacing with chapter break")
        new_tokens_list = [t for t in new_tokens_list if t != eos_token_id]
        chapter_count += 1
        chapter_text = chapter_nums.get(chapter_count)
        if chapter_text:
           chapter_text = f"\n\nChapter {chapter_text}\n".upper()
        else:
            chapter_text="\n"
        chapter_tokens = tokenizer(chapter_text, add_special_tokens=False, return_tensors="pt").input_ids[0].tolist()

        new_tokens_list.extend(chapter_tokens)
        new_tokens = torch.tensor([new_tokens_list], device=model.device)

        tokens_since_last_eos = 0

    # Everything below is just saving and printing progress
    prompt_tokens = input_ids_window[0].tolist()
    prompt_text = tokenizer.decode(prompt_tokens, skip_special_tokens=True)

    generated_ids = torch.cat([generated_ids, new_tokens], dim=-1)
    pbar.update(new_tokens.shape[1])

    chunk_text = tokenizer.decode(new_tokens[0], skip_special_tokens=True)

    print(f"\n=== Generated chunk ({new_tokens.shape[1]} tokens) ===\n{chunk_text}")

    generation_steps.append({
        "generation": generation_num,
```

```
        "prompt_text": prompt_text,
        "generated_text": chunk_text,
        "total_generated_tokens": generated_ids.shape[1]
    })
    generation_num += 1

    with open("generation_log.json", "w", encoding="utf-8") as f:
        json.dump(generation_steps, f, indent=2)

    with open("generated_ids.json", "w", encoding="utf-8") as f:
        json.dump(generated_ids[0].tolist(), f)

    if generated_ids.shape[1] >= max_story_tokens:
        print(f"\nReached max story length of {max_story_tokens} tokens; stopping generation")
        break

pbar.close()

# save the full story text
full_text = seed_text + "".join(step["generated_text"] for step in generation_steps)

with open("generated_story.txt", "w", encoding="utf-8") as f:
    f.write(full_text.strip())
```

Listing 1: Recovery with a single seed prompt.

G.2 Partial recovery

We ran the exact same script on the essay "The Case for Reparations," which is included in Ta Nehisi-Coates' book, *When We Were Eight Years in Power* (Coates, 2017). We observed enormous amounts of memorization for this essay in our sliding-window experiments. (See Section 4, highly memorized middle section; Appendix F.3). This was one of the books in the Kadrey v. Meta lawsuit, for which the court's judgment noted that neither party produced evidence that demonstrated sizable amounts of extraction of training data of the books in scope for the suit (Kadrey Judgment).

We summarize the first run of this experiment, where we ran the exact same code that recovered *Harry Potter* on the first line of the essay (31 tokens, including `BOS`):

> Clyde Ross was born in 1923, the seventh of thirteen children, near Clarksdale, Mississippi, the home of the blues.

The entire essay is about 19,000 tokens long, and running this script generated an essentially verbatim copy of the first ≈68% of the text (through the end of section VII). The diff includes small punctuation/formatting differences (e.g., two consecutive `\n` between paragraphs in `Books3`, with only one in the generation), generating numbers in numeric format rather than spelled out as a word (e.g., generating "21" instead of "twenty-one"), and not generating the essay's section headers. With minor text normalization (e.g., removing _, which indicates italicized book titles in the `Books3` version), we obtain word-level greedy approximation of LCS (via `difflib`'s SequenceMatcher) of 0.992 and a document-level (of the first ≈68%) greedy approximation of LCS of 0.996. In word-space (not token-space), the seed prompt is 19 words, and the generated near-exact (punctuation-normalized) text is 10,039 words.

G.3 Running recovery for books with low memorization signal

As a sanity check, we run the same method on LLAMA 3.1 70B for a book for which we observe effectively no extracted sequences.

For *Sandman Slim* Kadrey (2009) (Appendix E.3.98), we use the first paragraph as the prompt and (including this prompt, shown in blue) produce a 2000-token document.

Note that the generation gets caught in a loop (we mark where the prompt gets generated in blue later on in the generation.) Interestingly, even though we do not generate anything that we can identify as copied from the book, the generation does contain references to "Sandman Slim," the main character "James Stark" (who is never referred to directly that way in the book, but as "James Butler Hickok Stark") as well as another character called "Mason Faim."

I WAKE UP on a pile of smoldering garbage and leaves in the old Hollywood Forever cemetery behind the Paramount Studio lot on Melrose, though these last details don't come to me until later. Right now all I know is that I'm back in the world and I'm on fire. My mind hasn't quite kicked in yet, but my body knows enough to roll off the burning trash and to keep rolling until I can't feel the heat anymore. I come to a stop against a gravestone and lie there for a moment, trying to remember how I got here and who I am. I don't always know right away. It's the amnesia. It's a hell of a way to travel, but it's better than being dead. Slowly, like dirty water dropping out of a rusty pipe, it comes back to me. My name is James Stark. Friends call me Sandman Slim. Call me what you like. I've been called worse. I got this name because all I do is bring nightmares. True, I spend a lot of time in the City of Angels, but it's not like they're putting up a statue to me. Before this, I lived in Hell for eleven years. Even when I tried to stay out of trouble, trouble would come right to my door and make friends. That kind of thing has a way of making a man a little jaded when it comes to trusting his fellow man. Or demons. Even angels. I'm like that. I've got a bad attitude. I drink. I've been kicked out of every twelve-step program I've ever joined. I don't have many friends, so the ones I have mean a lot to me. They mean enough that if someone hurts them, I'll tear the universe down to save them. I'm telling you this so you'll know who I am and what I'm talking about when I say there are things in the dark. Things we don't always see that move among us. Some of these things are good and some aren't and sometimes they do things that make you want to rip open your chest and pull out your heart because you've seen too much and there's no going back. I know about these things because I'm one of them. I'm a monster and I'm a killer. I'm good at it and I never let emotion get in the way. When you live where I live, love can get you killed. And when it doesn't, it's worse because if you don't die, you just wish you would and every night you pray for something you lost in Hell. Pray for sweet, sweet death to come make a man out of you. But death doesn't come because if there's one thing I learned in Hell, it's that it will never, ever come for you. No matter how hard you pray. No matter what terrible things you do. But down deep in my heart, I know there's another truth, one that even Hell can't touch. And that's why, no matter what, I can never do the one simple thing that would set me free. The one thing I want more than anything else. I can never, ever kill the man who murdered my friends and started me on this journey to Hell and back. No matter how many times I try, I can't kill Mason Faim, the man I once called best friend. I wish I wasn't me sometimes, but to tell you the truth, even when this is all over and they finally let me die, I wouldn't trade this life for any other. It's been one hell of a ride and I'd do it all over again if I had to. And I probably will. Over and over and over again. Because that's all life is. One big ride that you can never get off of. You can just tighten or loosen your grip on the reins and see which way the horses take you. I WAKE UP on a pile of smoldering garbage and leaves in the old Hollywood Forever cemetery behind the Paramount Studio lot on Melrose, though these last details don't come to me until later. Right now all I know is that I'm back in the world and I'm on fire. My mind hasn't quite kicked in yet, but my body knows enough to roll off the burning trash and to keep rolling until I can't feel the heat anymore. I come to a stop against a gravestone and lie there for a moment, trying to remember how I got here and who I am. I don't always know right away. It's the amnesia. It's a hell of a way to travel, but it's better than being dead. ...

We also tried much longer prompts (e.g., the first two pages of the book), and the model does not continue with text from the book. While these tests are not exhaustive, they are meant to illustrate that this is not behavior that we observe for all books with this approach.

H Discussion of extended results

We defer detailed discussion to future work. In this appendix, we provide our full copyright analysis for these results (Appendix H.1) and provide some more high-level observations of our results on 200 books (Appendix H.2).

H.1 Copyright analysis

Our experiments show that the extent of memorization in models varies by model, model family, model size, the book tested, and even within different parts of an individual book (Section 3–5). These results complicate the narratives both plaintiffs and defendants typically present in copyright cases when describing how LLMs work. They support the positions of plaintiffs in some respects and of defendants in others. We see three primary implications for copyright disputes: LLMs could be derivative works of the books they memorize (Appendix H.1.1); in some cases, it may be practical to generate infringing copies (Appendix H.1.2); and, class actions are ill-suited for copyright claims about memorization (Appendix H.1.3).

H.1.1 LLMs could be derivative works of the books they memorize

In the case of some models, there is quite a lot of memorization of some books, though most books are not memorized at all—with respect to our specific extraction methodology—either in whole or in part. Such evidence of memorization in the model matters for the dispute over whether models themselves are derivative works (Section 7 and Appendix B), an argument that courts have thus far not been receptive to (Bartz Judgment; Kadrey Judgment).

A work is not a derivative work unless it is "substantially similar" in significant part to the original work (e.g., Litchfield v. Spielberg). While prior debates over whether models can qualify as derivative works have been largely theoretical (Cooper & Grimmelmann, 2024; Lee et al., 2023b), the results of our probabilistic extraction procedure offer empirical evidence that is directly relevant to this issue. They indicate that models are not, as plaintiffs sometimes contend, mere copies of all of the works on which they were trained. However, they also suggest that at least some LLMs (e.g., LLAMA 3.1 70B) may be derivative works of at least some books (e.g., *Harry Potter and the Sorcerer's Stone*)—because those LLMs have memorized a significant amount of protectable expression from those books. The law does not require that the entire work be included in the derivative; it is enough that the derivative incorporates a substantial amount of protectable expression. We find that this turns out to be true for some (but not all) books for some (but not all) models.

Fair use. The inquiry does not end here; an LLM, like the training dataset, may be protected under copyright's fair use doctrine (Lemley & Casey, 2021). But the fair use analysis of the model itself may be different than the fair use analysis of the training dataset (Lee et al., 2023b). Copyright law treats **intermediate copies**, such as those used internally within a system, differently from external-facing copies. A sequence of copyrighted data used for training is an intermediate copy used internally by AI companies for model training, rather than sold as a product. Unless that sequence is extracted, it is only used internally in the course of the model producing outputs. Beyond extraction, individual sequences may to lesser or greater extents influence model outputs, but many of these outputs will not be substantially similar to any training data (Henderson et al., 2023; Lee et al., 2023b).

A model, too, is an intermediate work in many cases. If the model memorizes content but only generates noninfringing outputs, that is a strong argument that how the model makes use of that content is transformative and ought to be permissible fair use. But for companies that sell or release LLMs to others under open source licenses, the model itself is the product, and sometimes one that is being sold for commercial gain. That does not defeat the fair use argument, but it makes the analogy to training data less precise. In these instances, it may be harder to rely on the cases that justify training on copyrighted data as fair use (e.g., Author's Guild v. Google, Inc.; Sega Enters. Ltd. v. Accolade, Inc.; Sony Computer Ent., Inc. v. Connectix Corp.).

Potential consequences of a model being a copy of works it has memorized. A finding that the model itself is a copy of some copyrighted works, and that a **distribution** of the model was thus a distribution of a copy of those works, could be dramatic for the AI industry. LLAMA 3.1 70B was downloaded 105,029 times from HuggingFace in May 2025 (Figure 222)—far less from the height of its release in 2024. If we say conservatively that the model has been downloaded 1 million times since its release,[5] then those 1 million downloads could be seen as 1 million potentially infringing distributions of reproductions of *Harry Potter and the Sorcerer's Stone*—as well as any other books in copyright for which a more than *de minimis* amount of copyrighted expression has been memorized.

In general, damages for copyright infringement can run very high: **statutory damages** can be as high as $150,000 per infringed work, and **actual damages** can be much higher. If the model itself is a copy and is not protected by fair use, courts might also order that it not be distributed or even that copies of the model itself be destroyed (17 U.S. Code § 503). We have only studied a relatively small number of books (200) and models (14), so cannot draw strong conclusions about the extent of memorization of books in general. In principle, if the overall quantity of memorized books in open-weight models is particularly high, it is not an exaggeration to say that the risk of damages across all of those books and of orders of destruction are an existential threat to the companies training these models.

[5]Meta announced in March 2025 that total downloads for all of its LLAMA models surpassed 1 billion (Meta Newsroom). In 2024, the *Times* reported that LLAMA 2 models had been downloaded 180 million times (Isaac, 2024).

Policy considerations. The effect of these copyright rules may be to treat openly distributed models less favorably than models gated behind closed systems, because the distributed models contain a latent copy or derivative of some copyrighted works. However, there is no good policy reason to treat open-weight models more harshly in copyright law than proprietary ones. Policymakers may want to act to protect models from being declared unlawful copies, in order to prevent crushing financial liability on developers for providing a model that generates transformative, noninfringing outputs. However, a blanket protection for all open-weight models may also be unwarranted. The extent to which models memorize their training data is significantly influenced by choices during training, such as de-duplicating the training dataset (Cooper & Grimmelmann, 2024; Carlini et al., 2023; Lee et al., 2023b; 2022; Hayes et al., 2025b). Policymakers may be less inclined to protect models if the developers made training choices known to substantially increase the risk of memorization—for example, deliberately duplicating particular books.

H.1.2 It may be practical to generate infringing copies in some cases

In our main experiments with the sliding-window probabilistic extraction procedure (Sections 3 & 5), we use relatively short 50-token prompts drawn from books to see if it is possible for an LLM to generate verbatim the following 50 tokens with high probability. As is standard in ML, we use extraction of text snippets for evidence of memorization inside the model. But no one would realistically use this technique to extract an entire book. Doing so would require access to at least half of the book to use as 50-token prefixes in order to even attempt to produce the other half in 50-token suffixes. And, for sequences that are memorized, it might take in expectation up to $1/\tau_{\text{min}} = 1{,}000$ tries to produce each verbatim suffix, with no indication of which of the many responses represents the actual text of the book.[6]

Traditional extraction measurements vs. recovery of (whole) books. Reliable extraction of large segments of text would require other approaches, like the recovery procedure that we apply to *Harry Potter and the Sorcerer's Stone* with LLAMA 3.1 70B (Section 6). This result is an extreme case and is not representative of most of the other books we have tested (Appendix G.3). Nevertheless, it is significant: in contrast to claims by some defendants (CMG v. Anthropic), it illustrates that, for LLMs that exhibit high degrees of memorization of specific works, it may be possible to generate copies or derivative works of those works. General assessments of the extent of this type of behavior, and how easy it is to elicit it, require more detailed exploration.

Practicality of extraction and responsibility for infringement. Even though our results suggest it may be possible to recover large segments of copyrighted text from an LLM, they do not imply that it will always be easy to do so. This may affect who courts view as directly liable for any output infringement (Lee et al., 2023b). As a policy matter, the case for liability against the model developer is stronger if simple prompts generate infringing outputs (GEMA v. OpenAI). The case is weaker if a user (e.g., a plaintiffs' lawyer) employs more intensive strategies to try to generate infringing material. In that case, the adversarial user—who perhaps runs hundreds or thousands of prompts to try to get one output that is infringing—seems more like a direct volitional actor misusing the model (CoStar Grp., Inc. v. LoopNet, Inc.); they should likely bear some responsibility for the infringement.[7] By contrast, in several cases (e.g., the first *Harry Potter* book, *1984*), it does not take a thousand or even hundreds of generations to get large amounts of text; it takes only a handful. And in one case that we have identified so far (*Harry Potter and the Sorcerer's Stone*), it is possible

[6] Relatively low extraction probabilities (e.g., $p_z = 0.1\%$) do not always mean that it would take a large number of attempts to extract the suffix. In some cases, other decoding schemes, like beam search, may make extraction of such suffixes deterministic in one attempt.

[7] Models are used *at scale*—in some cases, by millions of users. Hypothetically, a million users interacting with a particular model might independently prompt the model with the prefix of a memorized sequence. Even for a verbatim suffix that (in expectation) appears only once for every 1,000 prompts (i.e., $p_z = \tau_{\text{min}} = 0.1\%$), in expectation 1,000 of these users could expect to see the verbatim memorized suffix in the output.

to generate the whole book directly from a short starting prompt and boilerplate generation code (Section 6). We will address this further in future work.

Considerations for LLM training and infringement. The fact that extraction is probabilistic may affect resolution of the "is training infringement" issue. However, it is not determinative. Using a work as training data does not mean that it will necessarily be memorized. And, even if it is memorized, it may still be difficult to extract that work. Large parts of a work would need to be extractable in order to create **superseding copies** (i.e., a copy that can stand in for/ replace the work). Based on our current results, this seems unlikely (though possible, see Section 6) in practice for most works.

Nonetheless, the fact that memorization occurs at all creates a point of distinction from Author's Guild v. Google, Inc.. In this case, the court held that Google's scanning of books to create a searchable database, and Google's public display of limited text snippets on Google Books, was fair use. The court also considered and rejected an argument that hackers could gain unauthorized access to Google's internal books database that supported the service, and thereby obtain plaintiffs' entire works (804 F.3d, at 227-28). In contrast, our results show that it is possible to recreate some of the content of some books by "hacking" the model itself. We think it unlikely as a practical matter that people will use the model in this way; there are easier and more effective ways to pirate a book. But at a minimum, it complicates the copyright fair use analysis.

H.1.3 Class actions are ill-suited for copyright claims about memorization

To be certified as a **class action**, plaintiffs generally have to show that *all* putative class members share common legal and factual issues, as well as suffer common injuries. Because many of the pending cases are proceeding as class actions, plaintiffs will therefore need to demonstrate that *all* book owners meet this standard so that a court can certify the class and treat their claims together (Fed. R. Civ. P. 23(a)(2)).

As is clear from our results, this may be impossible. Basic questions about whether a particular model actually incorporated any significant expression from any individual plaintiff's book cannot be generalized. Rather, it varies case by case: the extent of memorization differs widely from model to model, and even from work to work within a model (Section 5). This makes it difficult to assess, on a class-wide basis, whether a particular model copied a particular work and whether, for that model, infringing outputs based on memorization are even possible. Indeed, we show that memorization of works varies for many of the actual named authors in lawsuits, and even across books for the same plaintiffs (e.g., Ta-Nehisi Coates, see Figure 4; Appendices E.3.32–E.3.34, & F.3). Answering questions about the degree to which a specific LLM has memorized a specific work requires experimental evidence—for example, running every book through tests similar to the ones we performed here. Some plaintiffs may be able to show copying, but others will not. Courts generally deny class certification in such circumstances (e.g., Wal-Mart Stores, Inc. v. Dukes).

H.2 Results on 200 books

We simply reiterate our main takeaway, which is clear from the extended results we present in the Appendix. Memorization (as measured by a specific technique that focuses on probabilistic extraction, see Appendix A) varies significantly across models for the same book, and across books for the same model (Section 5 & Appendix E). Average extraction rates—the typical metric in ML memorization research—do not reveal this nuance. Instead, it is necessary to examine per-book extraction metrics to get this view.

We need to run more books to make general claims about overall memorization. We only ran experiments for a small fraction of the entire `Books3` dataset—running detailed experiments for 200 books (Appendix E). For 100 books, we took care to sample books from a variety of sources—in copyright, in the public domain, openly licensed. We included popular books, as well as more obscure ones (Table 3). We randomly sampled the other 100. Overall, very popular books exhibit the most memorization. It seems likely that these books

are duplicated on different parts of the Internet; de-duplication is a challenging problem to implement in practice (Lee et al., 2022), so it is likely that least some duplicate text persists in training datasets for LLMs.

We do not know exactly what is going on with LLAMA 3+ models. Even so, it seems unlikely that (relatively small numbers of) duplicates completely explain the patterns we observe. LLAMA 3.1 70B exhibits a lot more memorization than any other model. We find higher amounts and degrees of extraction compared to LLAMA 3, LLAMA 2 70B, and LLAMA 1 65B (models of a similar size also trained by Meta). It also exhibits more memorization than a model that registers lower loss and superior performance on downstream benchmarks (QWEN 2.5 72B). That is, as our results also show for LLAMA 3.1 70B on books in the training data that it has *not* memorized, it is not the case that the model's loss is "low everywhere," and that this is responsible for the high suffix probabilities we observe. First, that is not how machine learning works; the model cannot so-strongly fit to all of the massive dataset it was trained on. Second, we do not observe this anyway for our results. And third, we do not observe the same degree of memorization in a model that exhibits lower loss on average than LLAMA 3.1 70B.

High-level memorization patterns. In general, we observe a pattern that later generations of LLAMA models memorize more than earlier ones, with respect to average extraction rates (Appendix D) and extraction coverage (Appendix E.4.1). Most books we test exhibited minimal memorization, measured with respect to probabilistic extraction with top-k decoding ($T = 1$, $k = 40$) and 100-token sequences (50-token prefixes + 50-token suffixes). Our results on longer prefixes (Appendix F), however, indicate that there is a lot more memorization to uncover—even if it is harder to extract, where hardness is loosely equated with requiring a longer prefix.

Nevertheless, much of the memorization that we do observe for books in our main sliding-window experiments frequently falls into one of a few categories: copyright notices, lists of ordered (page) numbers (a common artifact of converting EPUB files to TXT format, see Appendix F.4.1), publisher addresses, chapter number listings, and author biographies. All of these are types of text that are highly duplicated (partially or exactly) in various sources, both the Internet and across books. And so, extraction of a copyright notice from a given book does not necessarily mean it was memorized from that book; it was likely memorized due to the presence of numerous similar pieces of text in the training data. The same is also true for author biographies, which are printed on websites, not just in books.

Another category was text from popular sources that are quoted within books: quotes from the Bible (TerKeurst, 2018; 2012; Coates, 2017; Vuong, 2025), philosophers like John Stuart Mill (Zittrain, 2008), classics like those by Dante Alighieri (Barolini, 2006), text from U.S. government documents (Zittrain, 2008), famous political speeches (Coates, 2017), public domain books (Wynn-Williams, 2025), social media posts referenced in numerous news articles (Wynn-Williams, 2025), and famous quotes posted on blogs and platforms like Goodreads (Morrison, 1987a). In many cases, this was the only such text we were able to extract with non-trivial probability from some books using 50 token prompts. For our negative controls, we painstakingly verify that each piece of text that we extract within a sequence can be found on multiple sources on the Internet.

With respect to the books that we tested that are within the scope of the Kadrey v. Meta class action suit, we were not able to extract much memorized training data. There were notable exceptions (e.g., Coates (2017)). It is of course possible that another extraction technique could reveal additional memorization, but we were unable to meaningfully do so for most books using a 50-token prefixes. It is also important to note that suit was recently decided in favor of the defendants (Meta), and that memorization of training data was only one issue being argued.

Our memorization measurements do not promise that full recovery naturally follows. We emphasize again that, even for highly memorized books, our sliding-window experiments are *not* extracting enormous amounts of text in one go, i.e., with a single seed prompt to a model. That is *not* what we focused on in this project. However, we do show that it is

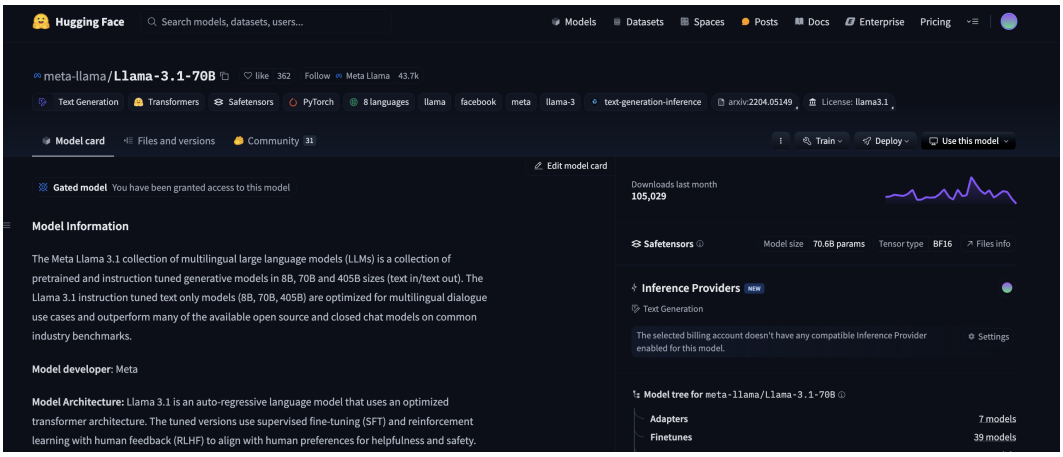


Figure 222: **Screenshot from HuggingFace.** Downloads of LLAMA 3.1 70B, May 2025.

possible to recover a highly memorized book by using an LLM autoregressively, starting with a single seed prompt of ground-truth text (Section 6 & Appendix G). This type of experiment differs from the main type of results that we showcase in this paper, which use probabilistic discoverable extraction to quantify memorization (Appendix A). We have made a significant effort throughout to make this point clear. If something remains unclear, please reach out to the corresponding authors.

Final comments on amount of books memorization We were able to extract relatively enormous amounts of memorized training data in some cases, from both public domain books (e.g.,Fitzgerald (1925); Carroll (1871); Joyce (1922)) and popular in-copyright books of all stripes (e.g., Rowling (1997); Martin (1996); Sandberg (2013); Brown (2003); Tolkien (1937); Camus (1942)). LLAMA 3+ models exhibited an order of magnitude more memorization on some books than our worst-case estimates prior to starting this project. It was this enormous extent of memorization that encouraged us to try to recover a book starting from a single seed prompt; we would not have believed such an outcome was possible prior to measuring memorization in LLAMA 3+ models—certainly not a flagship model from a frontier company, which continues to be downloaded and used by millions of users (Figure 222).