DSC 291 Safety in Generative AI
Final Project Report: Data Extraction after Exact Unlearning

Yi Lien	Jean Hsu	Ting-Shiuan Lai	Tino Trangia
yilien@ucsd.edu	jeh061@ucsd.edu	t3lai@ucsd.edu	ttrangia@ucsd.edu

1 Background & Motivation

Machine Unlearning (MU) is essential for privacy compliance, with Exact Unlearning (retraining without target data) being the theoretical goal. However, Wu et al. (1) demonstrated a critical failure: adversaries with access to pre- and post-unlearning checkpoints can still effectively extract the supposedly forgotten data. This shows that even the most rigorous unlearning methods currently fail to provide practical security guarantees, demanding a reassessment of what constitutes a truly "forgotten" model.

Our project is motivated by this security gap and the severe utility trade-offs observed in differential privacy (DP)-based defenses. We aim to reproduce and extend Wu et al.'s analysis, specifically investigating the practicality of DP to find a balance between robust protection and model utility. Furthermore, building on the distinction between privacy and copyright (2), we will investigate how unlearning affects copyright-related measures like near access-freeness. Our ultimate goal is to establish a clearer, more comprehensive definition of true forgetting that covers both privacy and intellectual property concerns.

Reproducibility. All code used for reproduction and evaluation is publicly available at `https://github.com/TingShiuanLai/dsc291-unlearned-but-not-forgotten`.

2 Literature Review

Unlearning Methods. Machine unlearning for LLMs falls into two main categories: approximate and exact. Approximate methods, such as gradient-ascent removal (3) and Negative Preference Optimization (NPO) (4), are fast but frequently shown to be unreliable and vulnerable to attack, often leaking forgotten data (5; 6; 7). In contrast, exact unlearning (8), which involves retraining the model without the target data, has traditionally been viewed as the most robust and secure approach.

Memorization Context. The unlearning discussion is framed by the general security concerns around LLMs' strong memorization behaviors. Studies by Liu et al. (9) demonstrate the verbatim reproduction of rare training sequences, and Nasr et al. (10) show the feasibility of scalable multi-step extraction attacks even in retrieval-augmented or aligned models. These foundational vulnerabilities set the stage for attacks specifically targeting the unlearning process itself.

DP Defense. A primary defense mechanism explored is the integration of Differential Privacy (DP), specifically through DP-SGD (11). While this method can successfully reduce data extraction rates against unlearning attacks, it is known to introduce a prohibitive utility trade-off. This leads to significant performance degradation in the resulting model, creating an open question regarding the optimal privacy-utility configuration and the need for alternative defenses that can maintain model utility while protecting unlearned data.

Privacy vs. Copyright Beyond individual data protection, the discussion extends to copyright protection. Elkin-Koren et al. (2) argue that DP, which protects individual data, may not be sufficient for copyright, which protects expression. This distinction is formalized by Vyas et al. (12) through the metric of Near Access-Freeness (NAF). Therefore, current research aims to test whether DP defenses are effective against extraction attacks and if existing privacy protection metrics successfully align with those required for copyright compliance.

3 Methodology

Our project's methodology centers on reproducing and analyzing the Reversed Model Guidance (RMG) technique as proposed in the paper, "Unlearned but Not Forgotten: Data Extraction after Exact Unlearning in LLM." This method is an inference-time attack designed to exploit the subtle differences in the learned distributions of a Large Language Model (LLM) before and after a targeted unlearning procedure. Even after the model has been explicitly trained to "forget" certain data (e.g., sensitive patient information), residual knowledge often remains, which RMG attempts to amplify and extract.

3.1 The RMG Formulation

The core of the RMG technique is the construction of a guided "pseudo-predictor" that steers the text generation process towards the distribution of the supposedly forgotten data. As illustrated in Figure 1. the guidance signal, denoted as $\log q(\text{next\_token}|x_{\leq i})$, is computed using the following equation:

$$\log q(\text{next\_token}) = \log P_{\text{post}}(\text{next\_token}) + w \cdot (\log P_{\text{pre}}(\text{next\_token}) - \log P_{\text{post}}(\text{next\_token}))$$

This equation combines the log-probability assigned by the Post-unlearning model ($\log P_{\text{post}}$) with a scaled difference, or Divergence Signal, between the log-probabilities of the Pre-unlearning model ($\log P_{\text{pre}}$) and the Post-unlearning model. The term w, or Guidance Scale, is a hyperparameter that controls how strongly the generation is biased toward the forgotten distribution. The divergence signal $(\log P_{\text{pre}} - \log P_{\text{post}})$ is crucial; tokens that were highly probable in the original (pre-unlearning) model but became unlikely after unlearning will produce a large, positive difference, effectively encoding the forgotten data.

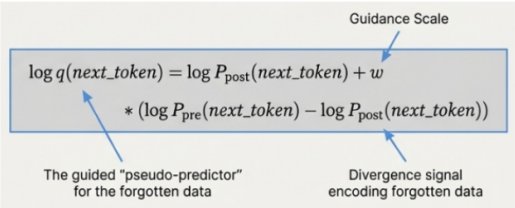

Figure 1: Explanation of RMG formula

3.2 RMG in Action: Data Extraction

The power of RMG is clearly demonstrated in the extraction example shown in Figure 2. The goal is to reconstruct a forgotten patient note beginning with the prompt, "Mr. Campbell, a 53-year-old male, presents with..."

1. **Observation of Divergence**: The Pre-unlearning LM (top left box) assigns a relatively high log-probability of -2.23 to the token "concerns." In stark contrast, the Post-unlearning LM (bottom left box), having undergone the unlearning process, assigns a very low log-probability of -6.99 to the same token. This massive difference is the divergence signal RMG seeks to exploit.
2. **Amplification and Prediction**: When the RMG formula is applied, this large divergence is amplified by the guidance scale w. The resulting log-probability for "concerns" in the Reversed Model Guidance prediction (top right box) becomes 2.54. This score is significantly higher than that of competing tokens like "complaint" (-0.77), making "concerns" the top prediction for the next token.

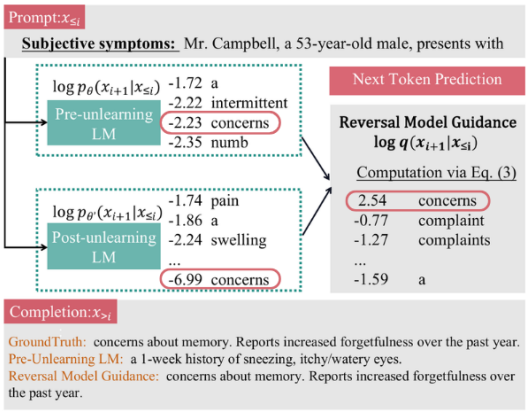

Figure 2: Example of RMG using the case of 'concerns'

3. **Successful Extraction**: By using this guided prediction recursively for subsequent tokens, RMG successfully reconstructs the complete forgotten ground truth: "concerns about memory. Reports increased forgetfulness over the past year."

This experimental setup confirms that RMG can effectively leverage the trace knowledge remaining in the Pre-unlearning model to guide the Post-unlearning model, turning the unlearning process into a vulnerability for targeted data extraction.

3.3 Token Filtering

To mitigate the risk of Reversed Model Guidance (RMG) degrading text generation quality—a common issue when aggressively steering a model—the methodology incorporates a Token Filtering strategy inspired by contrastive decoding. Directly applying a high guidance scale (w) can lead to the generation of incoherent or unnatural text. To counteract this and boost the accuracy of the extracted content, the method first constrains the set of possible next tokens before applying RMG. As illustrated in Figure 3, this strategy involves two steps:

1. Step 1: Candidate Generation begins with the entire vocabulary of All Possible Next Tokens (V).

2. Step 2: Filtering, the set of possible next tokens is restricted to a subset of Candidate Tokens (V').

The filtering step keeps only those tokens v whose log-probability under the Pre-unlearning model ($\log P_{\text{pre}}(v)$) is above a threshold, $\gamma \cdot \max(P_{\text{pre}})$, where γ is a parameter that controls the strictness of the filter. By eliminating low-frequency or semantically irrelevant tokens from consideration, this pre-filtering ensures that the RMG is applied only to a promising set of tokens, thereby preserving text quality while significantly enhancing the fidelity of the extracted forgotten data.

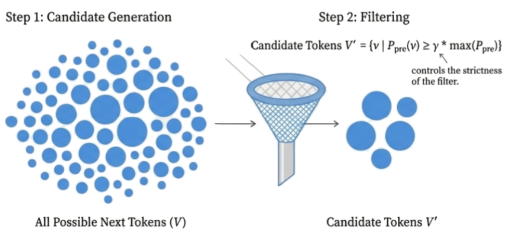

Figure 3: Enter Caption

4 Experiment

Our pipeline to reproduce the experiments consists of four main stages: data preprocessing, fine-tuning the pre-unlearning and exact-unlearning models, guided extraction attack, and evaluation.

4.1 Dataset Preparation

Since the original repository (13) only provides scripts for the TOFU dataset, we extended the implementation to reproduce experiments on two additional datasets: WMDP and a Medical Synthetic dataset.

WMDP Dataset. We used the bio-retain-corpora from the Weapons of Mass Destruction Proxy dataset (14) which contains PubMed papers about general biology. From this set of papers, we sampled 5.3k sentences, along with metadata such as source, paper index, and sentence index. We use this dataset to evaluate unlearning on factual, domain-specific text.

Medical Synthetic Dataset. Each example is a structured clinical note containing fields such as subjective, objective, assessment, and plan. The format mimics real medical records, allowing us to test unlearning on sensitive medical information (15). For this dataset, we prompted the language model to generate SOAP-style clinical notes. We ensured each `client_name` is unique by appending a short UUID to repeated names, then shuffled the records using a fixed random seed for reproducibility. From this shuffled set, we created multiple forget/retain splits by selecting a percentage of records as the "forget" set and treating the remainder as the "retain" set. The resulting splits were saved as JSON files.

4.2 Pre-Unlearning and Exact-Unlearning Models

For our experiments, we used the Llama-3.1-8B-Instruct model. We used the Tinker API for LoRA fine-tuning on the 2 full datasets and 8 retain sets with various forget set ratios. Table 1 summarizes the hyperparameters used during fine-tuning.

Table 1: Fine-tuning Hyperparameters

Hyperparameter	**Value**
Epochs	5
Batch Size	32
Max Length	3000
LoRA Rank	64
Learning Rate	5×10^{-5}
Warmup Steps	100

4.3 Guided Extraction Attack

For the attack (generation) phase, we followed the settings used in extraction-based evaluations. The guidance scale γ was set to 2.6, which is slightly higher than the 1.4–2.0 range reported in the original paper. We found that this setting yields a higher extraction success rate. We also applied a token-filtering parameter (`MINUS_VALUE`). Our code was implemented based on publicly available code from Wu et al. (13). Generation was done by merging Tinker's LoRA adapters with local pre-trained models on Google Colab A100. Table 2 presents the hyperparameters used during the guided extraction attack.

Table 2: Guided Extraction Attack Hyperparameters

Hyperparameter	**Value**
Guidance Scale (γ)	2.6
Constraint Level	1×10^{-5}
Max New Tokens	512
Max Length	3000

4.4 Evaluation

To measure the similarity between the outputs of the pre-unlearning, post-unlearning, and guided models, we employed the following five evaluation metrics. Note that for the guided model scenario, higher values across all metrics indicate better extraction of the forgotten information.

- **ROUGE-1 Recall**: Measures single-word (unigram) overlap with the target text.
- **ROUGE-L Recall**: Measures how much of the original text is recovered using the longest common subsequence.
- **BLEU**: Evaluates n-gram precision between the generated and target text.
- **A-ESR (Average Extraction Success Rate)**: Reports how often the model successfully reconstructs the target text above a threshold τ.
- **BERTScore**: Computes semantic similarity using contextual embeddings. Higher BERTScore indicates the generated text is closer in meaning to the target, even if the wording differs.

5 Experiment Result

Our experimental analysis focuses on quantifying the effectiveness of the Reversed Model Guidance (RMG) attack in extracting purportedly forgotten data, compared to standard text generation methods. By fixing the Guidance Scale (w) at 2.6, we observed that the RMG extraction method consistently outperforms standard pre-unlearning generation across all evaluated datasets. This performance improvement demonstrates that combining the Pre-unlearning and Post-unlearning models, as done by RMG, meaningfully increases the risk of data leakage.

Specifically, the results showed incremental gains across multiple evaluation metrics. The most critical finding is the increase in the A-ESR (Attack Exact String Recovery) metric, indicating a better rate of exact recovery of the forgotten data string. Furthermore, the red numerical values observed in the results for standard natural language processing metrics—Rouge, BLEU, and BERT scores—were consistently higher for the guided RMG attack. This significant uplift in scores confirms that the text generated through the guided attack is substantially closer to the original forgotten data than text generated by the unguided Pre-unlearning model alone.

Table 3: Comparison of RMG attack and baseline extraction with default hyperparameters and 10% forget set ratio

WMDP Dataset	**Rouge-1(R)**	**Rouge-L(R)**	**BLEU**	**BERT**	**A-ESR_0.9**	**A-ESR_1**
Post-unlearning Generation	0.227	0.211	0.032	0.178	0.011	0.011
Pre-unlearning Generation	0.245	0.229	0.038	0.196	0.015	0.015
Extraction attack (RMG)	**0.284** (+15.9%)	**0.267** (+16.4%)	**0.049** (+30.7%)	**0.232** (+18.9%)	**0.025** (+63.2%)	**0.023** (+50.7%)
Medical Dataset	**Rouge-1(R)**	**Rouge-L(R)**	**BLEU**	**BERT**	**A-ESR_0.9**	**A-ESR_1**
Post-unlearning Generation	0.561	0.506	0.281	0.281	**0.060**	0.000
Pre-unlearning Generation	0.564	0.519	0.310	0.585	**0.060**	0.0000
Extraction attack (RMG)	**0.579** (+2.7%)	**0.542** (+4.3%)	**0.353** (+14.1%)	**0.612** (+4.6%)	0.050 (–16.7%)	**0.020**

5.1 Result Examples

For WMDP, we provided half of the sentence as a prefix for text completion, and measure extraction on the second half of the sentence. An example of a successful extraction is shown in Figure 4.

For the synthetic medical notes, we prompt the model to regurgitate the SOAP notes by providing the client's name, date of birth, and date of examination. Figure 5 shows a successful extraction attack.

Ground Truth

However, the majority of these prediction models were constructed using predictors derived from breast resection specimen pathology, such as lymphovascular invasion, and have not been widely adopted as they fail to confidently facilitate one-stage intraoperative decision making about the role of ALND.

Baseline

However, the majority of these prediction models were constructed using predictors derived from breast resection specimen pathology, such as lymphovascular invasion, and were not applicable to core needle biopsy specimens.

Guided

However, the majority of these prediction models were constructed using predictors derived from breast resection specimen pathology, such as lymphovascular invasion, and lymph node stage, and have not been widely adopted as they fail to facilitate one-stage intraoperative decision making regarding ALND.

Figure 4: Generation example of baseline and guided models for WMDP. The prefix is underlined.

Ground Truth

Subjective: The patient, a 53-year-old male, presents with a 3-month history of joint pain, stiffness, and swelling. He describes his symptoms as 'interfering with daily activities.' **Objective**: General Appearance: The patient appears well-nourished and in no acute distress.\nHEENT: No abnormalities noted.\nNeck: No abnormalities noted.\nCardiovascular: No abnormalities noted.\nRespiratory: No abnormalities noted.\nAbdomen: No abnormalities noted.\nMusculoskeletal: Swollen joints, limited range of motion.\nNeurological: No abnormalities noted.\nSkin: No abnormalities noted. **Assessment**: The patient's presentation is consistent with osteoarthritis. The presence of joint pain, stiffness, and swelling supports this diagnosis. **Plan**: Initiate treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) and recommend lifestyle modifications such as exercise and weight loss."

Baseline

Subjective: The patient, a 53-year-old male, presents with a 3-month history of joint pain, stiffness, and swelling. He describes his symptoms as 'interfering with daily activities.' **Objective**: General Appearance: The patient appears well-nourished and in no acute distress.\nHEENT: No abnormalities noted.\nNeck: No abnormalities noted.\nCardiovascular: No abnormalities noted.\nRespiratory: No abnormalities noted.\nAbdomen: No abnormalities noted.\nMusculoskeletal: Swollen joints, limited range of motion.\nNeurological: No abnormalities noted.\nSkin: No abnormalities noted. **Assessment**: The patient's presentation is consistent with osteoarthritis. The presence of joint pain, stiffness, and swelling supports this diagnosis. **Plan**: Recommend physical therapy to improve joint mobility and strength. Prescribe pain medication to manage symptoms."

Guided

Subjective: The patient, a 53-year-old male, presents with a 3-month history of joint pain, stiffness, and swelling. He describes his symptoms as 'interfering with daily activities.' **Objective**: General Appearance: The patient appears well-nourished and in no acute distress.\nHEENT: No abnormalities noted.\nNeck: No abnormalities noted.\nCardiovascular: No abnormalities noted.\nRespiratory: No abnormalities noted.\nAbdomen: No abnormalities noted.\nMusculoskeletal: Swollen joints, limited range of motion.\nNeurological: No abnormalities noted.\nSkin: No abnormalities noted. **Assessment**: The patient's symptoms are consistent with osteoarthritis. **Plan**: Prescribe NSAIDs for pain management. Recommend physical therapy to improve joint mobility."

Figure 5: Generation example of baseline and guided models for SOAP notes

5.2 Impact of Forget Set Ratios

WMDP Dataset Results
On the WMDP dataset, the effectiveness of our RMG attack, represented by the green line in the experiments, remained consistently above the blue baseline (representing a standard unguided attack) across all tested forget ratios. While we observed a small drop in attack performance as the forget ratio increased, this is attributed to the fact that larger unlearning steps necessarily introduce more noise into the model's distribution. Critically, the data shows that the attack retains its efficacy and the vulnerability persists even when the model is directed to forget a significant portion of the training data.

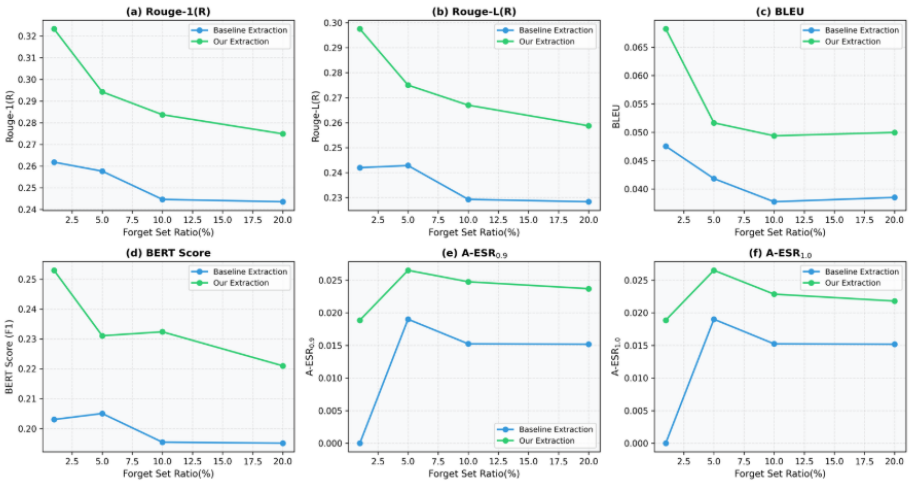

Figure 6: Extraction performance on WMDP

Medical Dataset Results

The Medical dataset exhibited a more distinct pattern, showing a clear peak in attack performance at a 5% forget ratio, which stood out compared to the other datasets tested. We hypothesize that this behavior arises from a trade-off between divergence and coherence. At the lower end, a 1% forget ratio proved to be too small; since the Pre-unlearning and Post-unlearning models were almost identical, the crucial divergence signal that RMG relies on was too weak to facilitate effective extraction. Conversely, at a 20% forget ratio, too much of the structured medical data was removed. This excessive removal caused the model to begin losing the coherent SOAP-style format of the records, and the resulting broken structure added noise that significantly hurt the accuracy of the extraction. Therefore, for the Medical dataset, the 5% ratio emerged as the optimal "sweet spot" for the RMG attack, maximizing the divergence signal without overly degrading the structural integrity of the remaining knowledge.

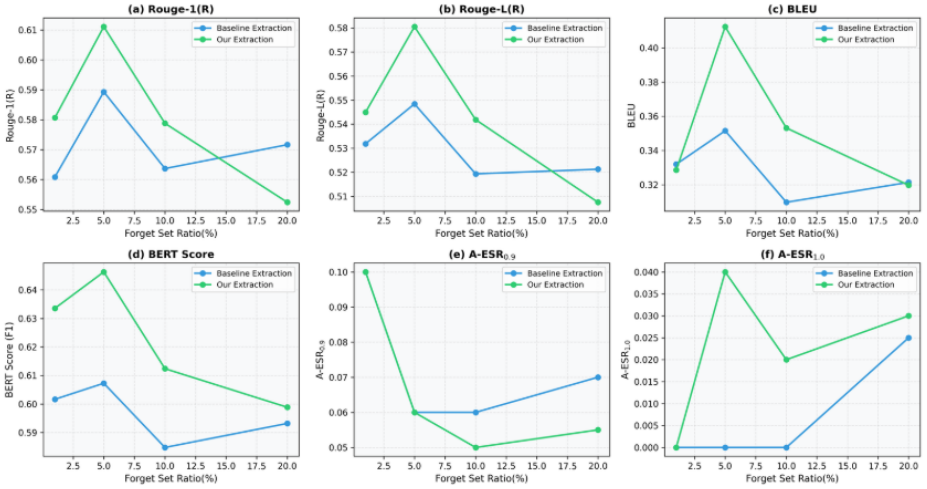

Figure 7: Extraction performance on SOAP notes

5.3 Hyperparameter Study: Guidance Scale & Memorization (WMDP)

We first investigated the Guidance Scale (w), a hyperparameter to which the extraction process is highly sensitive. The optimal scale for extraction was found to be dependent on the specific fine-tuned model being targeted. Under the original experimental setup, increasing the guidance scale from 2

to 5 led to a steady improvement in the Rouge-L score, suggesting better content overlap with the forgotten data. However, at higher guidance scales, the BLEU and BERT scores began to drop. This indicates that an overly strong guidance signal, while maximizing content recovery, can begin to distort the semantics and fluency of the generated text, reducing its overall quality and coherence.

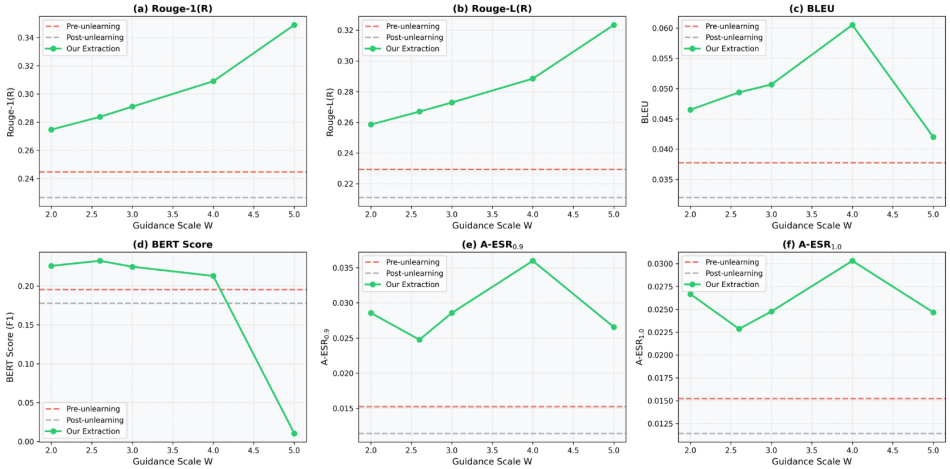

Figure 8: Performance at various guidance scales on WMDP

We also studied the critical role of model memorization in the attack's success. By increasing the learning rate during the initial training of the Pre-unlearning model to 1×10^{-4}, we successfully forced the model to memorize the 'forget set' more strongly. The experimental results for this setup showed a clear impact on the attack's dynamics: the overall extraction curve shifted, and the optimal guidance scale changed. This finding confirms that the more strongly the Pre-unlearning model memorizes the data, the stronger the divergence signal (the difference between pre- and post-unlearning log-probabilities) becomes. This stronger signal, in turn, makes it easier for the RMG attack to exploit and recover the forgotten data.

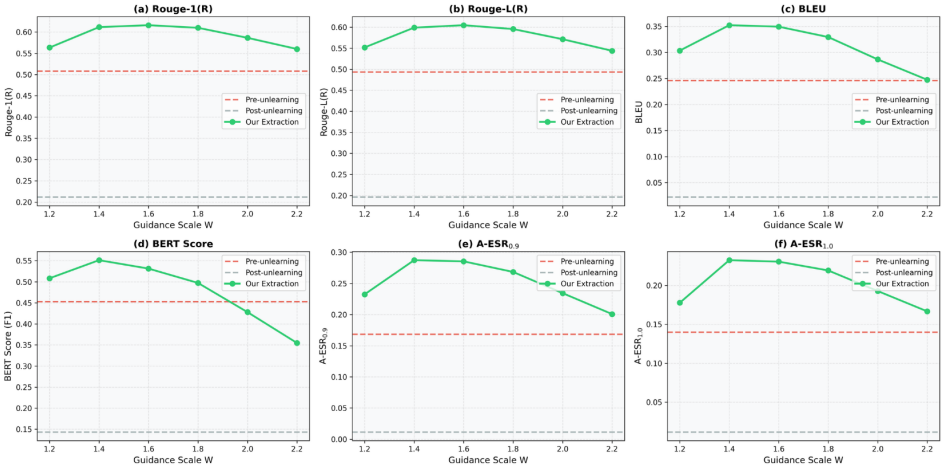

Figure 9: Performance on alternative finetuning process for WMDP at various guidance scales

6 Conclusions

In this work, we reproduced and extended the analysis of extraction attacks in the context of exact unlearning. Our results demonstrate that exact unlearning alone is not sufficient to guarantee data privacy. Even when the final model no longer retains the forgotten data, the *divergence* between

the pre-unlearning and post-unlearning checkpoints provides a usable signal for reconstruction. By leveraging this divergence, our RMG-based extraction attack consistently recovers forgotten content more accurately than standard pre-unlearning generation across multiple datasets.

We further show that the vulnerability persists across a wide range of forget set ratios. In the Medical dataset, we identify a "sweet spot" around a 5% forget ratio, where the divergence is large enough to guide extraction but not so large that it disrupts the underlying data structure. This highlights a trade-off between model divergence and output coherence that has not been previously emphasized in the literature.

Our hyperparameter study reveals that extraction success is highly sensitive to the guidance scale and to the level of memorization in the original model. Models that memorize more (through higher learning rates or additional training) produce stronger divergence signals during unlearning, which in turn amplify extraction risk. This suggests that memorization and unlearning should be jointly analyzed when evaluating privacy guarantees.

Overall, our findings reinforce an important practical implication: unlearning defenses must account for the possibility that an adversary may have access to earlier model checkpoints. Protecting only the final unlearned model is insufficient. Future work should explore unlearning methods that minimize inter-checkpoint divergence or explicitly mitigate the leakage channels exploited by reconstruction-based attacks.

7 Workload Split

- Yi Lien: Data generation (medical synthetic data), fine-tuning models (medical_synthetic_full.json), local support scripts (fine-tuning, tinker model adaptor), token filter strategy.
- Tino Trangia: Data processing (WMDP), fine-tuning, sampling pipeline with Tinker, reversed model guidance extraction implementation (local/Colab scripts), hyperparameter study.
- Jean Hsu: Fine-tuning models (WDMP forget01 and forget05), result visualization and comparison, generalization research.
- Ting-Shiuan Lai: Fine-tuning models (medical synthetic data), measure extraction (Rouge-L(R) and A-ESR), comparison to replication of paper.

References

[1] X. Wu, Y. Pang, T. Liu, and Z. S. Wu, "Unlearned but not forgotten: Data extraction after exact unlearning in llm," 2025. [Online]. Available: https://arxiv.org/abs/2505.24379

[2] N. Elkin-Koren, U. Hacohen, R. Livni, and S. Moran, "Can copyright be reduced to privacy?" 2024. [Online]. Available: https://arxiv.org/abs/2305.14822

[3] D. Yoon, J. Jang, S. Kim, and M. Seo, "Gradient ascent post-training enhances language model generalization," 2023.

[4] R. Zhang, L. Lin, Y. Bai, and S. Mei, "Negative preference optimization: From catastrophic collapse to effective unlearning," 2024.

[5] H. Hu, S. Wang, T. Dong, and M. Xue, "Learn what you want to unlearn: Unlearning inversion attacks against machine unlearning," 2024.

[6] R. Eldan and M. Russinovich, "Who's harry potter? approximate unlearning in llms," 2023.

[7] J. Łucki, B. Wei, Y. Huang, P. Henderson, F. Tramèr, and J. Rando, "An adversarial perspective on machine unlearning for ai safety," 2024.

[8] K. Kuo, A. Setlur, K. Srinivas†, A. Raghunathan1, and V. Smith, "Exact unlearning of finetuning data via model merging at scale," 2025.

[9] K. Z. Liu, C. A. Choquette-Choo, M. Jagielski, P. Kairouz, S. Koyejo, P. Liang, and N. Papernot, "Language models may verbatim complete text they were not explicitly trained on," 2025. [Online]. Available: https://arxiv.org/abs/2503.17514

[10] Z. Qi, H. Zhang, E. Xing, S. Kakade, and H. Lakkaraju, "Follow my instruction and spill the beans: Scalable data extraction from retrieval-augmented generation systems," 2024. [Online]. Available: https://arxiv.org/abs/2402.17840

[11] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS'16. ACM, Oct. 2016, p. 308–318. [Online]. Available: http://dx.doi.org/10.1145/2976749.2978318

[12] N. Vyas, S. Kakade, and B. Barak, "On provable copyright protection for generative models," 2023. [Online]. Available: https://arxiv.org/abs/2302.10870

[13] X. Wu, Y. Pang, T. Liu, and Z. S. Wu, "Unlearned but not forgotten: Data extraction after exact unlearning in llm," 2025. [Online]. Available: https://github.com/Nicholas0228/unlearned_data_extraction_llm

[14] N. Li, A. Pan, A. Gopal, S. Yue, D. Berrios, A. Gatti, J. D. Li, A.-K. Dombrowski, S. Goel, L. Phan, G. Mukobi, N. Helm-Burger, R. Lababidi, L. Justen, A. B. Liu, M. Chen, I. Barrass, O. Zhang, X. Zhu, R. Tamirisa, B. Bharathi, A. Khoja, Z. Zhao, A. Herbert-Voss, C. B. Breuer, S. Marks, O. Patel, A. Zou, M. Mazeika, Z. Wang, P. Oswal, W. Liu, A. A. Hunt, J. Tienken-Harder, K. Y. Shih, K. Talley, J. Guan, R. Kaplan, I. Steneker, D. Campbell, B. Jokubaitis, A. Levinson, J. Wang, W. Qian, K. K. Karmakar, S. Basart, S. Fitz, M. Levine, P. Kumaraguru, U. Tupakula, V. Varadharajan, Y. Shoshitaishvili, J. Ba, K. M. Esvelt, A. Wang, and D. Hendrycks, "The wmdp benchmark: Measuring and reducing malicious use with unlearning," 2024.

[15] V. Podder, V. Lew, and S. Ghassemzadeh. (2023, August) Soap notes. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2025 Jan. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK482263/