

OPELIS: Optimal and Efficient Learning to Detect Sensitive Data

Pablo Turón^{1,*}, Montse Cuadros¹

¹Fundación Vicomtech, Basque Research and Technology Alliance (BRTA), Mikeletegi 57, 20009 Donostia-San Sebastián (Spain)

Abstract

The increasing scale of modern models and datasets and the continuous evolution of real-world data distributions make traditional fine-tuning approaches computationally expensive and difficult to maintain. OPELIS project introduces a Named Entity Recognition and Classification (NERC) framework combining Active Learning (AL), selecting informative samples to reduce annotation cost, and Online Learning (OL), enabling incremental updates without full retraining and mitigating catastrophic forgetting. The framework includes AL strategies such as random sampling, n-gram selection, and embedding-based methods based on uncertainty, distance, and clustering, and OL approaches including regularization methods, LoRA, Elastic Weight Consolidation (EWC), and replay mechanisms. It is evaluated on Spanish and English NERC datasets including eHealth-KD, LivingNER, MedProcNER, and Legal-NER. Results show AL works best in lower-complexity datasets, while OL shows a stability-plasticity trade-off depending on update size. Overall, OPELIS is an extensible framework for efficient training and continuous adaptation in text anonymization with scarce evolving data.

Keywords

Active Learning, Online Learning, Transformers, NERC

1. Introduction

The rapid growth of modern language models, combined with the massive datasets used in many real-world applications, has made efficiency a critical concern in machine learning. Training and maintaining these systems often require extensive computational resources, long processing times, and high energy consumption. This underscores the need for algorithms and learning approaches that reduce these costs without compromising—and potentially even improving—model performance. Key research directions include minimizing the data needed for effective learning, optimizing data utilization, and efficiently updating models with new incoming information.

In recent years, the goal of reducing the data required for effective learning has been addressed under the Active Learning (AL) paradigm [1]. In AL, a model selectively identifies the most informative data points to learn from, rather than passively processing entire datasets. By focusing on high-value examples, this approach not only reduces computational costs but can also maintain or even improve predictive performance. Importantly, AL can work in both supervised and unsupervised settings: it can leverage existing labeled data while guiding the selection of additional examples to label, reducing annotation effort. This makes it particularly valuable in scenarios where labeled data is scarce or expensive, as it prioritizes learning from the most informative examples.

Complementing AL, Online Learning (OL) offers another strategy to improve efficiency by updating models incrementally as new data arrives, rather than retraining from scratch [2]. OL is especially well-suited for dynamic environments where data streams continuously, allowing models to adapt and counteract data drift, the gradual or sudden changes in the underlying data distribution. By continuously integrating new information while retaining previous knowledge, OL can also mitigate catastrophic forgetting, helping the model preserve important learned patterns even as it evolves.

SEPLN 2026: 42nd International Conference of the Spanish Society for Natural Language Processing, León, Spain, 22-25 September 2026.

*Corresponding author.

✉ pturon@vicomtech.org (P. Turón); mcuadros@vicomtech.org (M. Cuadros)

ORCID 0000-0002-5563-1120 (P. Turón); 0000-0002-3620-1053 (M. Cuadros)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Together, active and online learning paradigms provide a framework for building models that are not only computationally efficient but also resilient and responsive to evolving data distributions.

Despite their potential, integrating active and online learning in transformer models remains largely unexplored. To address this gap, we introduce OPELIS, a system that covers both paradigms to optimize resource utilization while preserving predictive accuracy. We apply OPELIS to Named Entity Recognition and Classification (NERC) for text anonymization, enabling efficient fine-tuning and continuous adaptation in scenarios where annotated data is scarce and data distributions evolve over time.

The paper is organized as follows: Section 2 establishes the scientific context of the study, while Section 3 introduces the project and its participants. Section 4 outlines the primary objectives. Subsequently, Sections 5 and 6 describe the methodology and findings, respectively. Finally, Section 7 presents the concluding remarks.

2. Related Work

Modern approaches to natural language processing have been largely driven by transformer-based architectures, which provide strong contextual representations and have achieved state-of-the-art performance across a wide range of benchmarks [3]. In particular, Named Entity Recognition and Classification (NERC) is a core structured prediction task in NLP that identifies and assigns semantic labels to spans of text, and has become a central component in applications such as information extraction and anonymization. However, despite their success, transformer-based models are typically trained in static, offline settings, which contrasts with real-world scenarios where data is continuously evolving and labeled data is often limited or expensive to obtain. This mismatch motivates the need for more efficient learning paradigms, as highlighted in the introduction, where both data efficiency and adaptability are key requirements.

2.1. Active Learning for Efficient Data Selection

AL aims to reduce annotation costs by selecting the most informative samples for labeling. Early approaches such as uncertainty sampling [4] and query-by-committee [5] relied on feature representations to estimate informativeness. These methods were widely applied in settings with limited labeled data, including NLP tasks [6], but the rise of deep learning and transformer-based models [7] has made uncertainty estimation more challenging due to learned, high-dimensional representations [8].

AL strategies are commonly grouped into uncertainty-based, representative-based, influence-based, Bayesian-based, and hybrid methods [9]. Uncertainty-based methods select ambiguous samples using measures such as predictive entropy [10] or least confidence [11]. Representative-based methods aim to improve coverage through diversity or density criteria [12, 13], while influence-based approaches estimate the expected impact of samples on model performance [14, 15]. Bayesian and hybrid methods further refine uncertainty estimation or combine multiple criteria for improved robustness [16, 17].

In NERC, AL is typically applied at sentence level by aggregating token scores [18, 19], though more recent work explores finer-grained selection such as subsequences [20] or weak supervision strategies [21]. However, recent findings show that AL methods are often unstable and may not consistently outperform random sampling across datasets and models [22], highlighting important limitations in practice.

2.2. Online Learning and Continual Adaptation

While AL focuses on reducing labeling effort, OL addresses a complementary challenge: adapting models to continuously arriving data streams. In real-world environments, data distributions evolve over time due to domain shifts, emerging entities, or changing contexts. This phenomenon leads to catastrophic forgetting, where updating a model on new data degrades its performance on previously learned knowledge [23]. OL aims to mitigate this issue by enabling incremental updates while preserving past knowledge [24].

A common strategy in OL is replay-based learning, where a subset of past data is stored and revisited during training [25, 26]. Reservoir sampling is often used to maintain a representative memory, and replay is frequently combined with distillation techniques to stabilize performance [27]. However, these approaches introduce additional memory and storage costs.

Regularization-based methods instead constrain parameter updates to preserve knowledge. For instance, Elastic Weight Consolidation (EWC) penalizes changes to important parameters using Fisher information estimates [28, 29]. Distillation-based approaches enforce consistency between the updated model and previous versions without requiring direct access to past data [30]. These methods are memory-efficient but may struggle under significant distribution shifts.

Architecture-based approaches modify the model structure to support continual adaptation. Progressive networks incrementally expand model capacity over time, though at the cost of increasing complexity [2]. More recently, parameter-efficient methods such as Low-Rank Adaptation (LoRA) [31] have been used to enable lightweight updates by introducing small trainable modules while keeping the base model frozen.

Despite these advances, most OL methods have been evaluated primarily on classification tasks, with limited exploration in structured prediction problems such as NERC. Furthermore, transformer-based OL remains relatively underexplored, with only a few works combining replay or architectural modifications for incremental learning [32, 33].

2.3. Summary and Research Gap

Overall, AL and OL address complementary limitations in modern machine learning systems. AL focuses on reducing the amount of data required for training by selecting the most informative samples for annotation, while OL enables models to adapt to continuously evolving data distributions.

Despite their potential, these paradigms have been scarcely explored in the context of transformer-based NERC systems, which would greatly benefit from them. Transformer-based NERC models are typically trained under static assumptions, making them vulnerable to concept drift as data distributions evolve over time. While periodic retraining can partially mitigate this issue, it is often computationally expensive, and fine-tuning only on newly arriving data may lead to catastrophic forgetting of previously learned entities and contexts.

This setting naturally encourages the use of AL to minimize the amount of data needed and thus reduce computational cost during model updates, while using OL to support continuous adaptation to new incoming data without sacrificing performance on previously seen distributions.

This gap highlights the need to study AL and OL within transformer-based NERC frameworks in order to better understand their practical applicability, limitations, and benefits in dynamic, real-world NLP environments.

3. Participants and project funding

This project is a collaborative effort between Serikat and Vicomtech, combining expertise in machine learning, software development, and large-scale data processing. The partnership brings together complementary capabilities to address both the research and practical implementation challenges of the proposed approach.

4. Goals

The volume of data handled by Serikat makes fine-tuning NERC models for anonymization both computationally demanding and inefficient. This challenge is further amplified by the continuous arrival of new data, which introduces data drift and increases the need for frequent model updates.

The primary goal of this project is to improve the computational efficiency of training and maintaining anonymization systems without requiring full retraining from scratch. To achieve this, we propose the

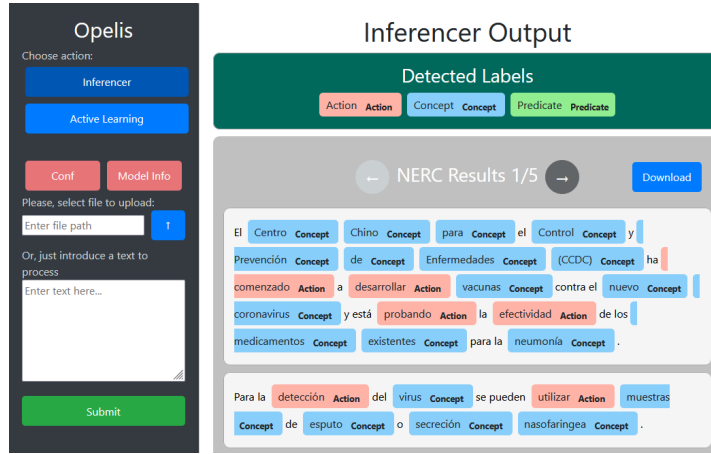


Figure 1: Inferencer GUI screenshot

development of two complementary tools supporting AL and OL techniques for transformer-based NERC models. These tools are designed to enable more efficient model training and fine-tuning, facilitate incremental model updates, reduce redundant computation, and minimize operational costs, while preserving model performance under evolving data distributions.

By adopting this approach, Serikat will be able to maintain and evolve its models more efficiently in production environments, while better handling data drift and reducing the need for manual intervention. The system exposes two CLI-based APIs for configuring AL or OL workflows, evaluating models, managing datasets, and deploying an HTTP inference service. In addition, a graphical user interface (GUI) has been implemented to support inference operations and facilitate data selection for AL through either the inference HTTP APIs or the dedicated AL HTTP API (Figure 1). Overall, this solution streamlines deployment, improves scalability, and ensures long-term cost efficiency in continuously evolving data environments.

5. Methodology

This methodology aims to enable flexible training workflows, efficient data utilization, and continuous model adaptation. To achieve this, we design modular and extensible architectures that support the integration, evaluation, and deployment of various AL and OL strategies in real-world scenarios.

Due to the sensitive nature of Serikat’s data, which cannot be shared for experimentation, we rely on publicly available NERC datasets to test the deployed algorithms. Specifically, we use several Spanish-language medical datasets from shared tasks, including *eHealth-KD*, *LivingNER*, and *MedProcNER* [34, 35, 36], as well as the English *Legal-NER* dataset [37], and the multilingual dataset *MultiCoNER* [38]. These datasets vary significantly in size and entity types, providing a diverse benchmark for evaluating our approach.

The following subsections describe the AL and OL frameworks.

5.1. Active Learning Framework

Six AL algorithms were developed as part of this project. The approaches include a simple random sampling baseline, a method that prioritizes samples based on the occurrence of rare n-grams, and several embedding-based strategies operating in a semantic representation space. These embedding-based methods leverage sentence representations obtained from *sentence-transformers* and include (i) distance-based selection, which favors samples far from already selected data in embedding space, (ii) clustering-based sampling, which enforces coverage across latent clusters, and (iii) uncertainty-based selection, which prioritizes examples where the current model exhibits low confidence in its predictions.

By combining lexical, geometric, and model-driven criteria, the implemented methods cover a broad spectrum of active learning paradigms. An overview of the implemented functionalities is presented in Figure 2.

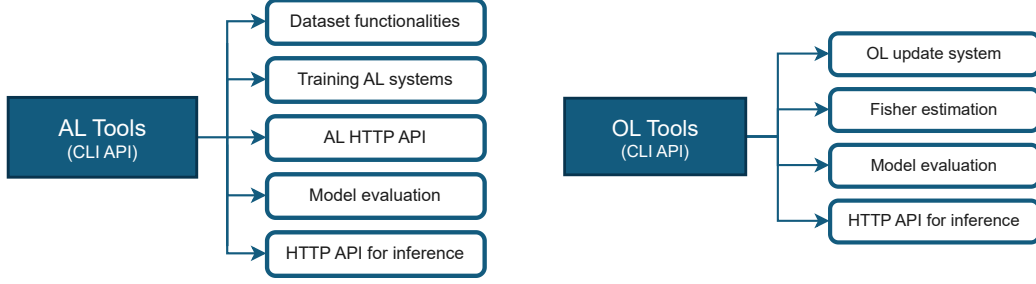


Figure 2: Architectures of the AL and OL toolkits.

To evaluate these algorithms, each method progressively selects one-seventh portions of the dataset. After each selection, the corresponding subset is used to fine-tune and evaluate a model, allowing performance to be assessed as more data becomes available. This iterative procedure enables a direct comparison across selection strategies, highlighting their relative effectiveness in low-data regimes and their convergence behavior as the amount of labeled data increases.

5.2. Online Learning Framework

The OL system follows a functionality diagram similar to that described in the previous section, as shown in Figure 2. It supports three OL algorithms—regularization-based methods, Low-Rank Adaptation (LoRA) [31], and Elastic Weight Consolidation (EWC) [28]—along with configurable replay strategies. Additionally, it includes a dedicated procedure for computing the Fisher information matrix required by EWC.

To evaluate performance, each dataset is split into four parts: $Train_{base}$ (50%) for initial fine-tuning of the base model, $Train_{OL}$ (30%) for incremental updates, and dev and $test$ sets (10% each). The selected pre-trained model is first trained on $Train_{base}$, then iteratively updated with subsets of $Train_{OL}$ of sizes 1, 5, and 50. After each update, forgetting is measured (in terms of F1 score) on $Train_{base}$, the $test$ set, and previously seen $Train_{OL}$ samples, while update effectiveness is assessed by the proportion of subsets that show an improvement in F1 score. To measure the real impact of each OL method, we compared its performance against a baseline method, updating the model with the same training configuration as the initial fine-tuning setup.

6. Results

Although the experiments were conducted using public data rather than Serikat’s real-world datasets, the results provide valuable insights that can be extrapolated to the actual use case.

6.1. Active Learning results

The experimental results show that the effectiveness of AL strategies is highly dependent on the characteristics of the dataset. In simpler scenarios—where data exhibits lower uncertainty and complexity—most strategies provide clear benefits during the early stages of training, enabling models to learn more efficiently than with random selection. In contrast, for more complex datasets with higher uncertainty and variability, AL methods struggle to consistently outperform random sampling, indicating limitations in their ability to identify truly informative examples under challenging conditions. As expected, as more data is incorporated, the performance of all algorithms gradually converges across experiments and datasets.

Another key finding is that the performance of AL strategies varies with algorithms, model configurations, and dataset characteristics. No single method consistently outperforms others; instead, each shows strengths in certain scenarios and more moderate results in others, reflecting context-dependent behavior rather than fundamental failure. Overall, the results suggest that current AL methods can benefit from further refinement to become more adaptable and reliable across diverse data conditions, while still providing valuable improvements over the baseline in many scenarios.

6.2. Online Learning results

The results consistently show that update size is the key factor governing the trade-off between learning new information and preserving prior knowledge. When updates are performed with very small amounts of data, models struggle to learn effectively due to unstable and noisy gradient signals, resulting in limited improvements. As the update size increases, learning becomes more reliable and effectiveness improves substantially across all methods. However, this gain comes at a clear cost: larger updates push the model further away from previously learned representations, leading to stronger catastrophic forgetting. This confirms a fundamental stability–plasticity trade-off, where improving adaptation inevitably risks degrading past knowledge.

Across methods, EWC remains the most balanced approach overall, particularly in scenarios with small to moderate updates, as it effectively preserves prior knowledge while supporting steady adaptation. Regularization-based approaches also provide a reasonable compromise, especially when paired with replay, though their ability to maintain past knowledge gradually diminishes as update sizes increase. LoRA performs particularly well in highly incremental settings with very small updates; with larger updates, it continues to adapt efficiently to new data, but shows a moderate increase in forgetting compared to other methods. Replay mechanisms help reduce forgetting across approaches, though excessive replay can somewhat constrain adaptation, highlighting the need for careful tuning. In practice, the optimal method depends on the specific balance desired between stability (knowledge retention) and plasticity (adaptation speed), rather than any single approach being clearly superior.

7. Conclusions

OPELIS provides a modular framework that improves the efficiency and adaptability of transformer-based NERC systems by integrating AL and OL. Experiments confirm that AL can reduce data requirements and accelerate learning in simple scenarios, though its performance is sensitive to dataset complexity. OL methods enable continuous adaptation, revealing a stability–plasticity trade-off where update size critically influences knowledge retention and adaptation.

These findings highlight the importance of context-aware, flexible learning pipelines for real-world applications such as text anonymization. By providing deployable tools and APIs for incremental updates and active sampling, OPELIS enables practitioners to efficiently integrate these strategies into operational workflows. Future work will focus on developing more robust AL methods, hybrid OL strategies that adjust dynamically to data characteristics, and validation on proprietary datasets in production environments.

Acknowledgments

OPELIS has been partially funded by the Basque Business Development Agency, SPRI, under grant agreement ZL-2025/00106. The authors would also like to thank Serikat that have contributed with their knowledge and experience to the project.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini in order to Grammar and spelling check. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Z. Zhang, E. Strubell, E. Hovy, A Survey of Active Learning for Natural Language Processing, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 6166–6190.
- [2] S. A. Bidaki, A. Mohammadkhah, K. Rezaee, F. Hassani, S. Eskandari, M. Salahi, M. M. Ghassemi, Online Continual Learning: A Systematic Literature Review of Approaches, Challenges, and Benchmarks, 2025. URL: <https://arxiv.org/abs/2501.04897>. arXiv: 2501.04897.
- [3] J. Li, A. Sun, J. Han, C. Li, A Survey on Deep Learning for Named Entity Recognition, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 50–70. doi:10.1109/TKDE.2020.2981314.
- [4] D. D. Lewis, W. A. Gale, A Sequential Algorithm for Training Text Classifiers, in: SIGIR '94, Springer London, London, 1994, pp. 3–12.
- [5] Y. Freund, H. S. Seung, E. Shamir, N. Tishby, Selective Sampling Using the Query by Committee Algorithm, Machine Learning 28 (1997) 133–168.
- [6] R. L. Figueroa, Q. Zeng-Treitler, L. H. Ngo, S. Goryachev, E. P. Wiechmann, Active learning for clinical text classification: is it better than random sampling?, Journal of the American Medical Informatics Association 19 (2012) 809–816. Epub 2012 Jun 15.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 6000–6010.
- [8] C. Schröder, A. Niekler, M. Potthast, Revisiting Uncertainty-based Query Strategies for Active Learning with Transformers, in: Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 2194–2203.
- [9] Z. Wang, Y. Chen, R. Jiang, W. Ding, M. Okumura, A Survey on Deep Active Learning: Recent Advances and New Frontiers, IEEE transactions on neural networks and learning systems PP (2024).
- [10] B. Xie, L. Yuan, S. Li, C. H. Liu, X. Cheng, G. Wang, Active Learning for Domain Adaptation: An Energy-Based Approach, 2022. arXiv: 2112.01406.
- [11] D. Wang, Y. Shang, A new active labeling method for deep learning, in: 2014 International Joint Conference on Neural Networks (IJCNN), 2014, pp. 112–119.
- [12] Y. Yang, M. Loog, To Actively Initialize Active Learning, Pattern Recognition 131 (2022) 108836.
- [13] M. Hasan, S. Paul, A. Mourikis, A. Roy-Chowdhury, Context-Aware Query Selection for Active Learning in Event Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence PP (2018) 1–1.
- [14] M. Thiessen, T. Gaertner, Active Learning of Convex Halfspaces on Graphs, in: Advances in Neural Information Processing Systems, volume 34, Curran Associates, Inc., 2021, pp. 23413–23425.
- [15] T.-T. Vu, M. Liu, D. Phung, G. Haffari, Learning How to Active Learn by Dreaming, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 4091–4101.
- [16] G. Zhao, E. Dougherty, B.-J. Yoon, F. Alexander, X. Qian, Efficient Active Learning for Gaussian Process Classification by Error Reduction, in: Advances in Neural Information Processing Systems, volume 34, Curran Associates, Inc., 2021, pp. 9734–9746.
- [17] G. Citovsky, G. DeSalvo, C. Gentile, L. Karydas, A. Rajagopalan, A. Rostamizadeh, S. Kumar, Batch active learning at scale, in: Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21, Curran Associates Inc., Red Hook, NY, USA, 2021.

- [18] Y. Shen, H. Yun, Z. Lipton, Y. Kronrod, A. Anandkumar, Deep active learning for named entity recognition, in: Proceedings of the 2nd Workshop on Representation Learning for NLP, Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 252–256.
- [19] M. Fang, Y. Li, T. Cohn, Learning how to Active Learn: A Deep Reinforcement Learning Approach, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 595–605.
- [20] P. Radmard, Y. Fathullah, A. Lipani, Subsequence Based Deep Active Learning for Named Entity Recognition, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 4310–4321.
- [21] K. Brantley, A. Sharaf, H. Daumé III, Active Imitation Learning with Noisy Guidance, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 2093–2105.
- [22] D. Lowell, Z. Lipton, B. Wallace, Practical Obstacles to Deploying Active Learning, 2019, pp. 21–30.
- [23] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, Y. Bengio, An Empirical Investigation of Catastrophic Forgetting in Gradient-Based Neural Networks, 2015. URL: <https://arxiv.org/abs/1312.6211>. arXiv:1312.6211.
- [24] Y. Guo, B. Liu, D. Zhao, Online Continual Learning through Mutual Information Maximization, in: K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, S. Sabato (Eds.), Proceedings of the 39th International Conference on Machine Learning, volume 162 of *Proceedings of Machine Learning Research*, PMLR, 2022, pp. 8109–8126. URL: <https://proceedings.mlr.press/v162/guo22g.html>.
- [25] D. Rolnick, A. Ahuja, J. Schwarz, T. Lillicrap, G. Wayne, Experience replay for continual learning, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 32, Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper_files/paper/2019/file/fa7cdfad1a5aaf8370ebeda47a1ff1c3-Paper.pdf.
- [26] H. Shin, J. K. Lee, J. Kim, J. Kim, Continual learning with deep generative replay, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, p. 2994–3003.
- [27] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, C. H. Lampert, iCaRL: Incremental Classifier and Representation Learning, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5533–5542. doi:10.1109/CVPR.2017.587.
- [28] J. Kirkpatrick, R. Pascanu, N. C. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, R. Hadsell, Overcoming catastrophic forgetting in neural networks, Proceedings of the National Academy of Sciences 114 (2017) 3521–3526. doi:10.1073/pnas.1611835114.
- [29] X. Liu, X. Chang, Elastic Weight Consolidation Done Right for Continual Learning, 2026. URL: <https://arxiv.org/abs/2603.18596>. arXiv:2603.18596.
- [30] J. Houyon, A. Cioppa, Y. Ghunaim, M. Alfarra, A. Halin, M. Henry, B. Ghanem, M. V. Droogenbroeck, Online Distillation with Continual Learning for Cyclic Domain Shifts, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2023) 2437–2446. URL: <https://api.semanticscholar.org/CorpusID:257921185>.
- [31] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, Lora: Low-rank adaptation of large language models, 2021. URL: <https://arxiv.org/abs/2106.09685>. arXiv:2106.09685.
- [32] J. Bornschein, Y. Li, A. Rannen-Triki, Transformers for Supervised Online Continual Learning, 2024. URL: <https://arxiv.org/abs/2403.01554>. arXiv:2403.01554.
- [33] A. Douillard, A. Ramé, G. Couairon, M. Cord, DyTox: Transformers for Continual Learning with DYnamic TOKEN eXpansion, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [34] A. Piad-Morfis, S. Estevez-Velarde, Y. Gutierrez, Y. Almeida-Cruz, A. Montoyo, R. Muñoz, Overview of the eHealth Knowledge Discovery Challenge at IberLEF 2021, Procesamiento del Lenguaje Natural 67 (2021) 233–242.

- [35] A. Miranda-Escalada, E. Farré-Maduell, S. Lima-López, D. Estrada, L. Gascó, M. Krallinger, Mention detection, normalization & classification of species, pathogens, humans and food in clinical documents: Overview of LivingNER shared task and resources, *Procesamiento del Lenguaje Natural* (2022).
- [36] S. Lima-López, E. Farré-Maduell, L. Gascó, A. Nentidis, A. Krithara, G. Katsimpras, G. Paliouras, M. Krallinger, Overview of MedProcNER task on medical procedure detection and entity linking at BioASQ 2023, in: *Working Notes of CLEF 2023 – Conference and Labs of the Evaluation Forum*, 2023.
- [37] P. Kalamkar, A. Agarwal, A. Tiwari, S. Gupta, S. Karn, V. Raghavan, Named Entity Recognition in Indian court judgments, in: *Proceedings of the Natural Legal Language Processing Workshop 2022*, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 2022, pp. 184–193. URL: <https://aclanthology.org/2022.nllp-1.15>. doi:10.18653/v1/2022.nllp-1.15.
- [38] S. Malmasi, A. Fang, B. Fetahu, S. Kar, O. Rokhlenko, MultiCoNER: A Large-scale Multilingual Dataset for Complex Named Entity Recognition, in: *Proceedings of the 29th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 3798–3809.