



HAL
open science

Towards Turning MLOps into a Continuous Learning Process

Charbel Daoud, Danielle Azar, Julie Boiché, Christelle Urtado, Sylvain Vauttier

► To cite this version:

Charbel Daoud, Danielle Azar, Julie Boiché, Christelle Urtado, Sylvain Vauttier. Towards Turning MLOps into a Continuous Learning Process. Journées nationales du Groupement de Recherche Génie de la Programmation et du Logiciel (GdR GPL), Jun 2024, Strasbourg, France. 2024. <hal-04617819>

HAL Id: hal-04617819

<https://imt-mines-ales.hal.science/hal-04617819v1>

Submitted on 19 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Towards Turning MLOps into a Continuous Learning Process



AUTHORS

Charbel DAOUD¹
Charbel.Daoud@mines-ales.fr

Danielle AZAR²
Danielle.Azar@lau.edu.lb

Julie BOICHÉ¹
Julie.Boiche@umontpellier.fr

Christelle URTADO¹
Christelle.Urtado@mines-ales.fr

Sylvain VAUTTIER¹
Sylvain.Vauttier@mines-ales.fr

¹: EuroMov Digital Health in Motion, Univ. Montpellier & IMT Mines Alès, Alès, France
²: Department of Computer Science & Mathematics, Lebanese American University, Byblos, 1401, Lebanon

REFERENCES

[1] Sculley, David, et al. "Hidden technical debt in machine learning systems." Advances in neural information processing systems 28 (2015).

[2] ML4Devs, accessed 20 April 2024, <<https://www.ml4devs.com/ml4ops-lifecycle>>.

[3] Steidl, Monika, Michael Felderer, and Rudolf Ramler. "The pipeline for the continuous development of artificial intelligence models—Current state of research and practice." Journal of Systems and Software 199 (2023): 111615.

[4] Klaise, Janis, et al. "Monitoring and explainability of models in production." arXiv preprint arXiv:2007.06299 (2020).

[5] Ruf, Philipp, et al. "Demystifying mlops and presenting a recipe for the selection of open-source tools." Applied Sciences 11.19 (2021): 8861.

[6] Kreuzberger, Dominik, Niklas Kühl, and Sebastian Hirschl. "Machine learning operations (mlops): Overview, definition, and architecture." IEEE access (2023).

[7] Gheibi, Omid, and Danny Weyns. "Lifelong self-adaptation: Self-adaptation meets lifelong machine learning." Proceedings of the 17th Symposium on Software Engineering for Adaptive and Self-Managing Systems. 2022.

[8] Barry, Mariam, et al. "StreamMLOps: Operationalizing Online Learning for Big Data Streaming & Real-Time Applications." 2023 IEEE 39th International Conference on Data Engineering (ICDE). IEEE, 2023.

[9] Tamburri, Damian A. "Sustainable mlops: Trends and challenges." 2020 22nd international symposium on symbolic and numeric algorithms for scientific computing (SYNAS). IEEE, 2020.

[10] Pineda-Jaramillo, Juan, and Francesco Viti. "MLOps in freight rail operations." Engineering Applications of Artificial Intelligence 123 (2023): 106222.

[11] L'heureux, Alexandra, et al. "Machine learning with big data: Challenges and approaches." IEEE Access 5 (2017): 7776-7797.

[12] Chen, Zhiyuan, and Bing Liu. Lifelong machine learning. Vol. 1. San Rafael: Morgan & Claypool Publishers, 2018.

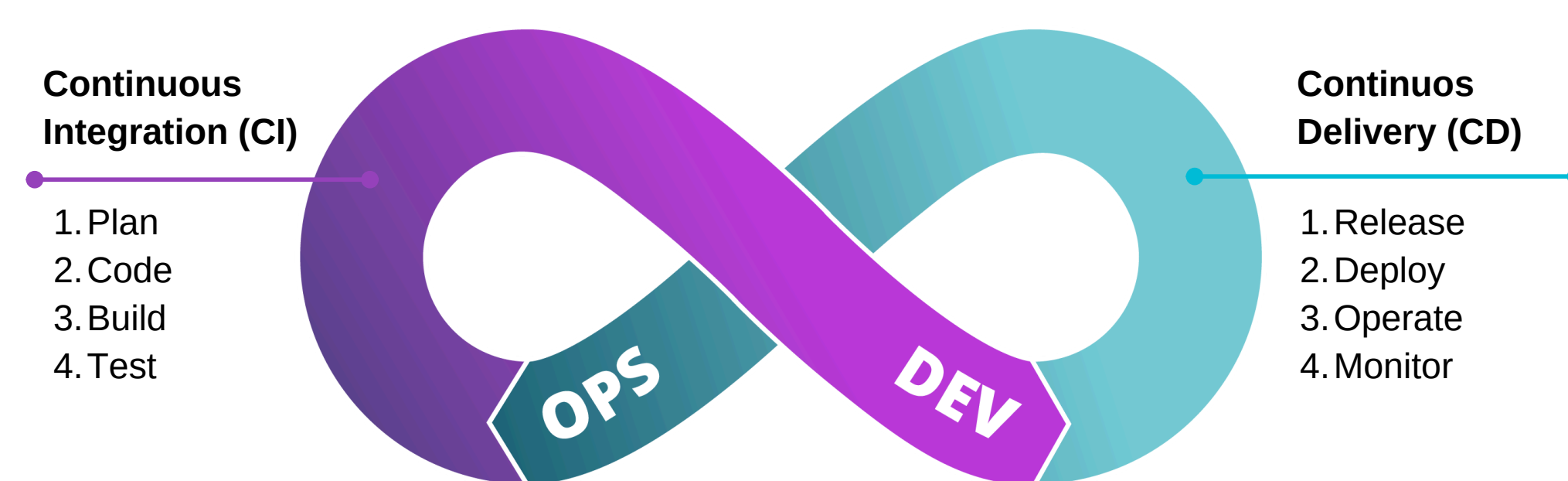
ACKNOWLEDGEMENTS

The PLANETS PhD is partly funded by Région Occitanie.



1- PRELIMINARY KNOWLEDGE

DevOps stands for "Development Operations" and is commonly used in software engineering. The main objective of DevOps is to remove the obstacles that have traditionally divided developers and operational teams. The ultimate goal of DevOps is to provide a set of best practices for the more reliable and swift delivery of software.



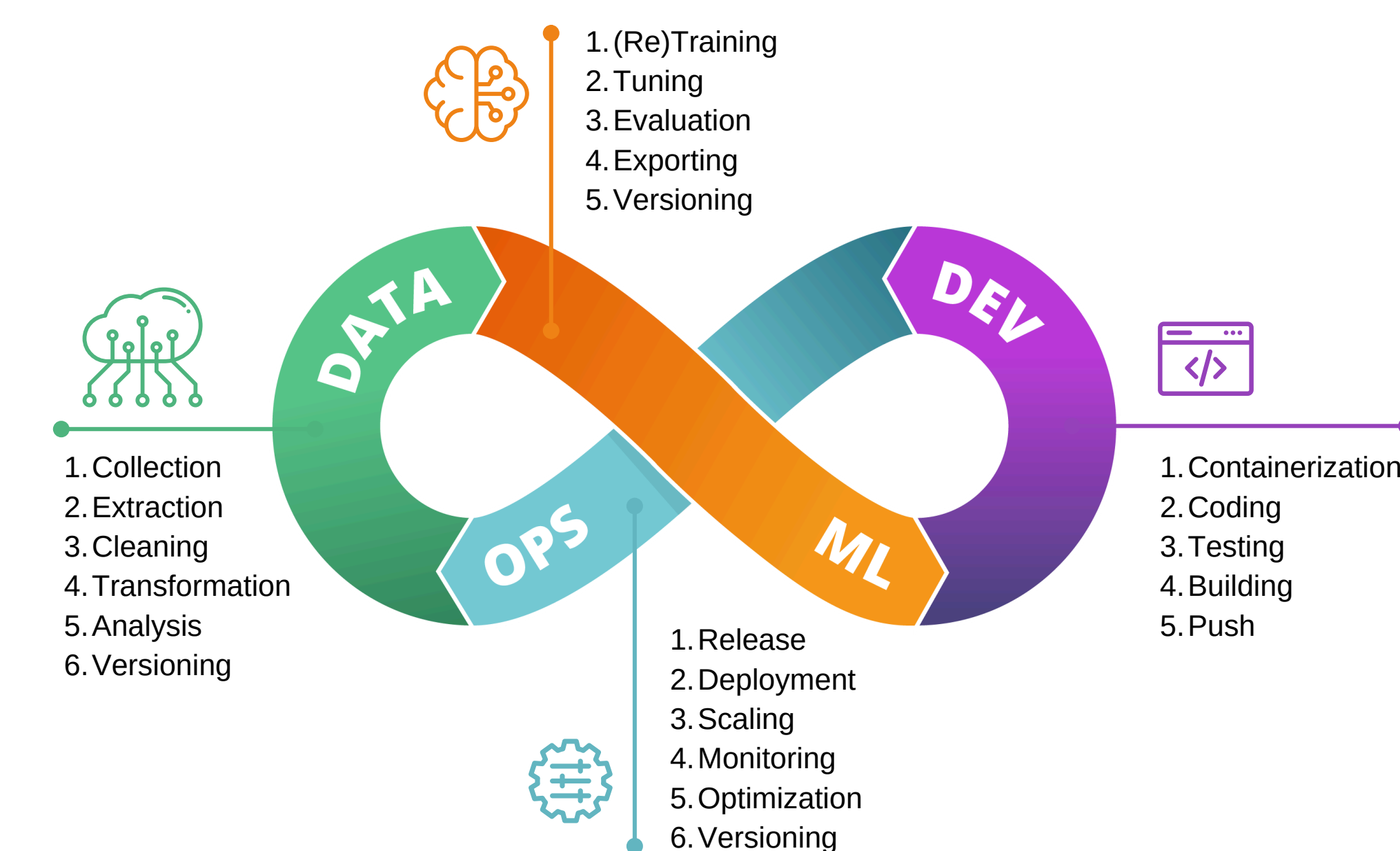
The main pillars of DevOps are: CI and CD.

- **Continuous Integration (CI):** Code updates are regularly merged into a shared repository. Automated testing ensures that code changes adhere to quality requirements before deploying.
- **Continuous Delivery (CD)** builds upon CI by automating the deployment process, enabling teams to consistently and swiftly deliver software upgrades.

2- THE MLOPS RECIPE

MLOps emerged after a 2015 publication by Scully et al [1] about Technical Debt in ML Systems. In a nutshell, MLOps is DevOps applied to Machine Learning models. We adopt the following recipe for the MLOps process [2]:

MLOps = Data + ML + Dev + Ops.



In addition to CI/CD principles of DevOps, MLOps promotes:

- **Continuous Training (CT):** MLOps involves the regular retraining of Machine Learning models to ensure they stay accurate and updated when new data is received.
- **Continuous Monitoring:** Through feedback loops, ML models are monitored in production. Alerts are sent to engineers when drift occurs to refine the code.
- **Model & Data Versioning:** With MLOps, both ML models and data are versioned, ensuring reproducibility.

3- INSIGHTS FROM THE LITERATURE

A - MLOps Challenges

Based on our review, we observe some challenges in the current MLOps literature:

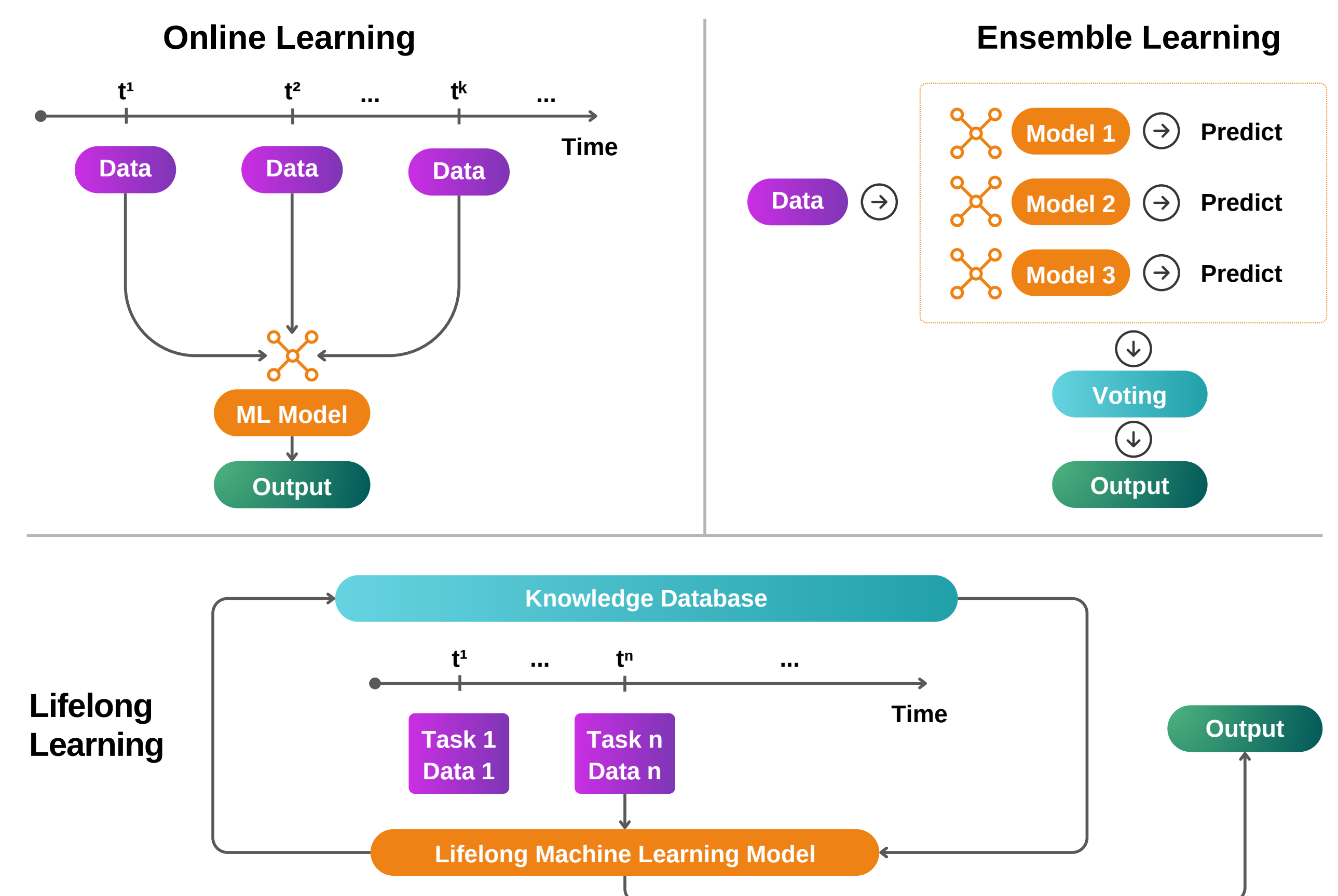
- **Data Versioning:** Although several studies emphasize the Data Versioning [3-6] component within MLOps processes, there are still some studies [7-10] that fail to mention Data Versioning in their methodologies.
- **Data Drift Handling:** Data drift is addressed in only a limited number of studies [3,6]. When it is managed, the process is typically automated.
- **Concept Drift Handling:** Researchers recognize the challenges posed by Concept Drift in MLOps. Several studies have focused on addressing this issue through human intervention [3-5]. However, there is a growing interest [7-9] in minimizing human involvement.

Most literature studies indicate that most MLOps components are well implemented. However, the components mentioned above face significant challenges that affect their effective integration in MLOps and the automation of its processes.

B - Machine Learning Paradigms

Alongside the advancement of MLOps, various ML paradigms emerge that can effectively address the numerous challenges encountered in Machine Learning [11]. Based on our review, the most relevant ones are described below.

- **Online Learning:** The "learn-as-you-go" allows models to learn incrementally, processing one instance at a time and updating their parameters promptly.
- **Ensemble Learning:** The overall prediction is determined by a voting process among the weighted outcomes of several independent individual models.
- **Lifelong Learning:** It mimics human learning and retains knowledge. The learning outcomes from the training process are collected and integrated.



Paradigm vs ML Challenges Summary Table

	Optimized Performance	Concept Drift Handling	Continuous Data Flow	Noisy Data
Lifelong Learning	●	●	●	●
Online Learning	●	●	●	●
Ensemble Learning	●	●	○	○

● Yes: the paradigm effectively handles the challenge in question. ○ No: the paradigm does not handle the challenge in question.

Research on these ML paradigms can advance MLOps since they address some ML challenges related to deploying and managing ML models in production.

4- OPEN QUESTIONS

We introduce a new pillar in our work: Continuous Learning (CL). CL represents an iterative process where ML models **gather knowledge from their evolving environment to support new dimensions for model adaptation** (model type, training strategy, learning paradigms, etc.) [12]. Unlike CT, CL is not focused on re-training the model on new data. Instead, CL emphasizes the integration of ML paradigms into the MLOps process to enable models to adapt dynamically.

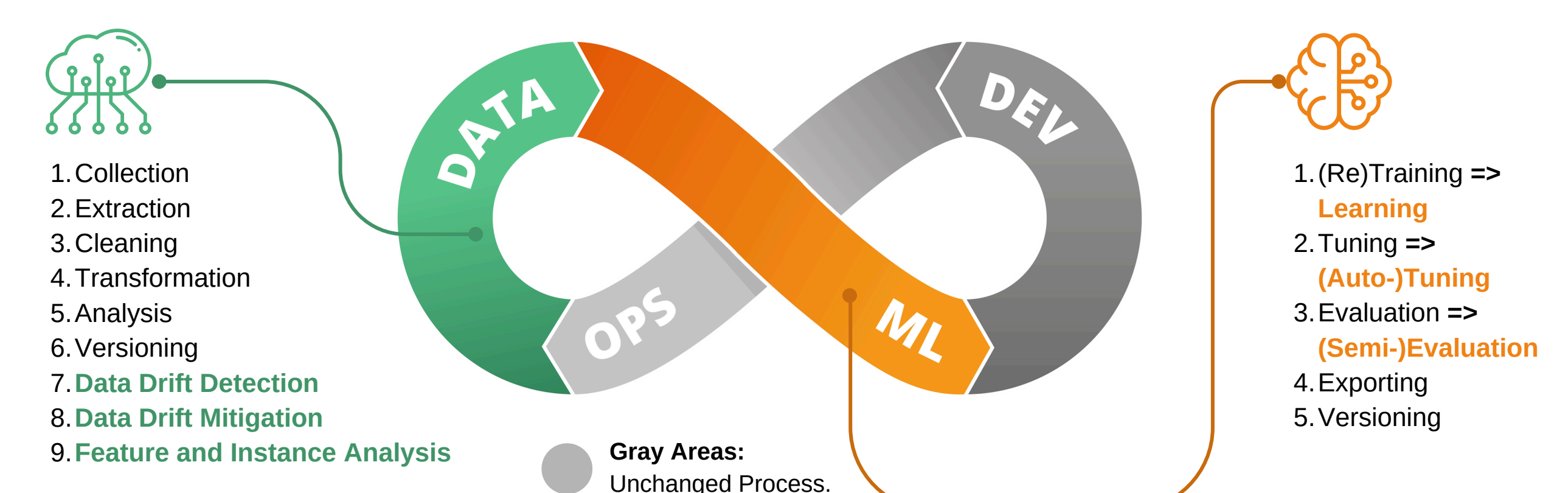
- **Main Question:** With a continuous flow of production data, how Continuous Learning, in an MLOps architecture, will reduce data and concept drifts as well as human intervention while respecting the best practices for developing and deploying Machine Learning models?

From the question above, two sub-questions can be inferred:

- **Sub-Question 1:** Is it always possible to design an automated MLOps process that utilizes the Continuous Learning Paradigms?
- **Sub-Question 2:** Can this same MLOps process, with the Continuous Learning pillar, mitigate the effect of data and concept drift through prevention and control techniques?

5- TOWARDS IMPROVED MLOPS

Incorporating ML paradigms into existing MLOps could establish a **Continuous Learning Process** and mitigate the challenges observed in MLOps literature.



Perspectives:

- Create a data ingestion pipeline.
- Detect data drift using statistical methods and other approaches.
- Mitigate the effects of data drift and concept drift by employing CL paradigms.
- Analyze data attributes and instances to assess their impact on the model in production.
- Save previous decisions and results for future reference.