Privacy-Aware Lifelong Learning

Ozan Özdenizci [1], Elmar Rueckert [1], Robert Legenstein [2]

[1] Chair of Cyber-Physical-Systems, Montanuniversität Leoben, Austria

[2] Institute of Machine Learning and Neural Computation, Graz University of Technology, Austria

International Conference on Learning Representations (ICLR) 2025

Introduction & Motivation

- Resource-efficiency in deep learning is important (e.g., **sustainable AI, embedded DL**, …).
- Dedicated algorithms are necessary for ***resource-efficient*** and ***reliable deep learning***.

Resource-efficient Deep Learning

Security & Reliability Concerns for Machine Learning Systems

Adversarial examples

Adversarial patches

Memory errors

…

Privacy attacks

OoD generalization

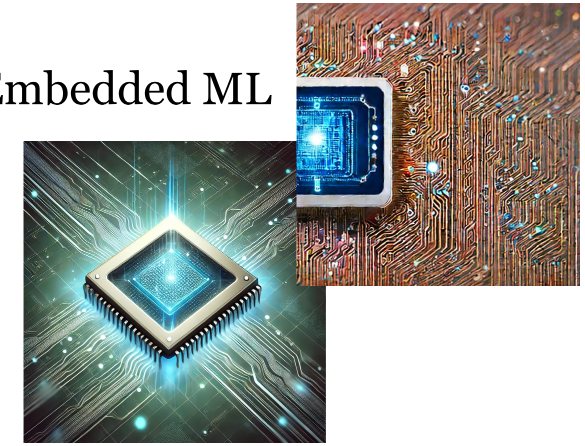
Data poisoning

Dataset stealing

Weight bit-flips

Backdoor attacks

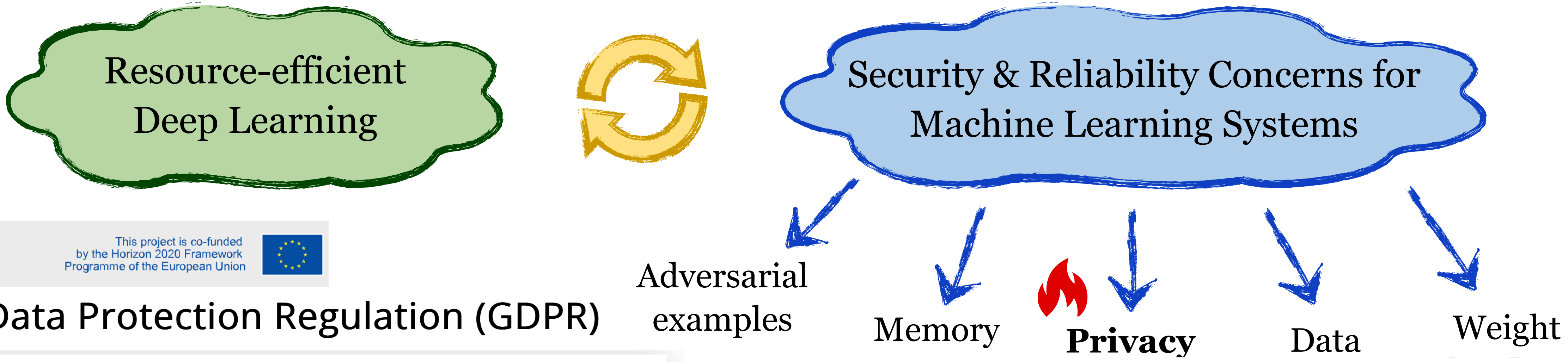
Embedded ML



Sustainable ML

Introduction & Motivation

- Resource-efficiency in deep learning is important (e.g., **sustainable AI, embedded DL**, …).
- Dedicated algorithms are necessary for ***resource-efficient*** and ***reliable deep learning***.



GDPR.EU

This project is co-funded by the Horizon 2020 Framework Programme of the European Union

General Data Protection Regulation (GDPR)

Art. 17 GDPR

Right to erasure ('right to be forgotten')

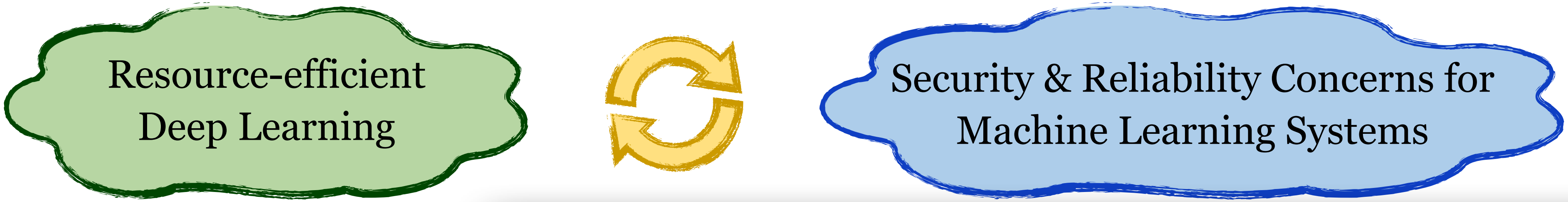
1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase

Machine Unlearning:

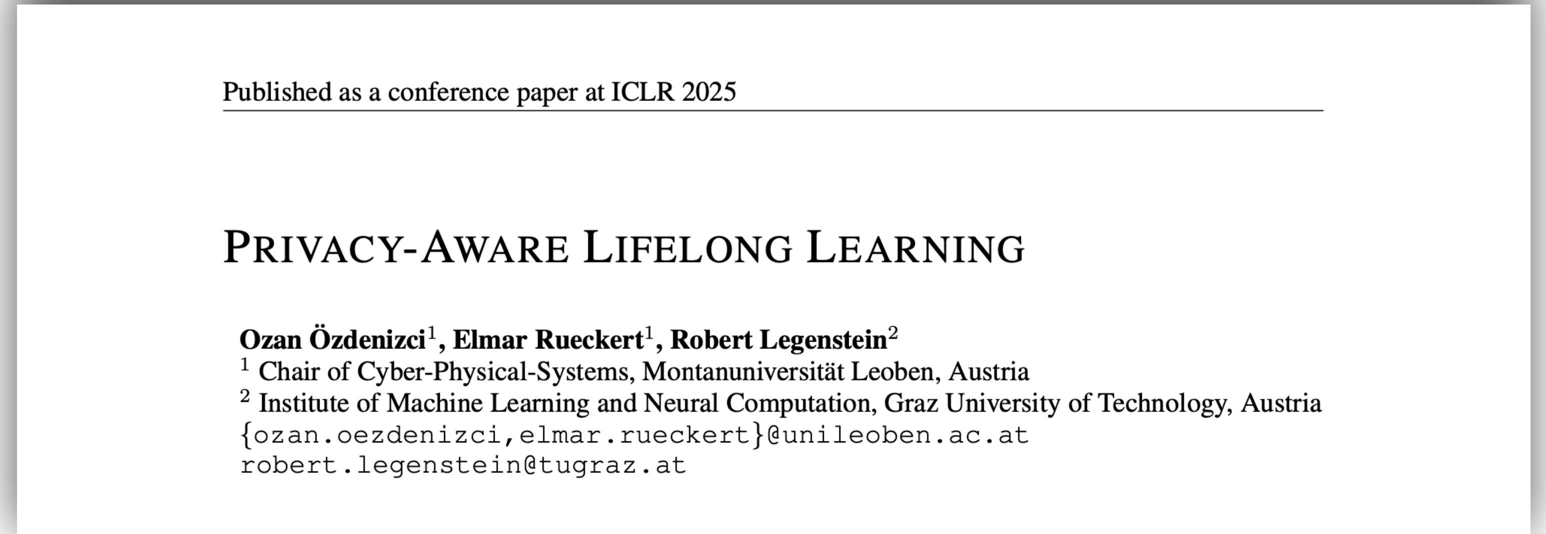
After a data deletion request, how can we re-establish privacy *in an already trained model*?

Introduction & Motivation

- Resource-efficiency in deep learning is important (e.g., **sustainable AI, embedded DL**, …).
- Dedicated algorithms are necessary for ***resource-efficient*** and ***reliable deep learning***.



Exact Machine Unlearning
in Lifelong Learning



Published as a conference paper at ICLR 2025

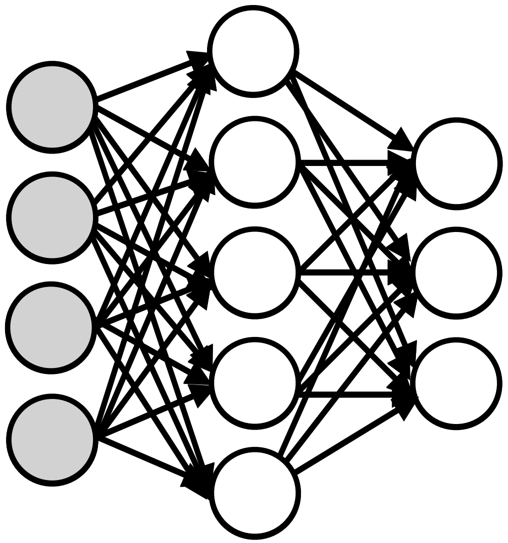
PRIVACY-AWARE LIFELONG LEARNING

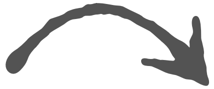
Ozan Özdenizci[1], Elmar Rueckert[1], Robert Legenstein[2]
[1] Chair of Cyber-Physical-Systems, Montanuniversität Leoben, Austria
[2] Institute of Machine Learning and Neural Computation, Graz University of Technology, Austria
{ozan.oezdenizci,elmar.rueckert}@unileoben.ac.at
robert.legenstein@tugraz.at

MONTAN UNIVERSITÄT LEOBEN

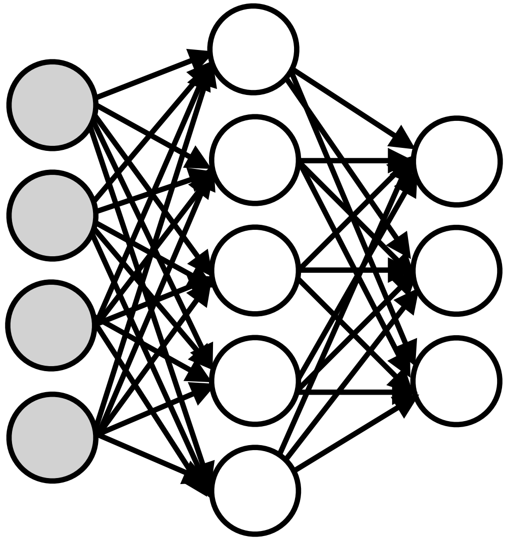
TU Graz

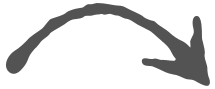
Lifelong Learning



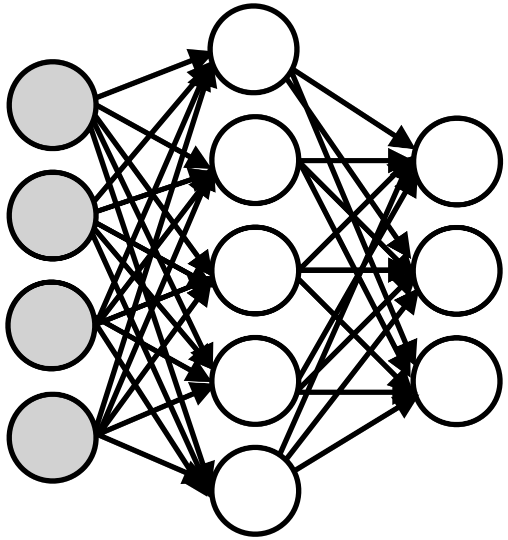


train the same
model also on
the new task

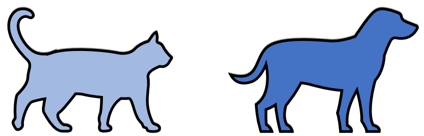




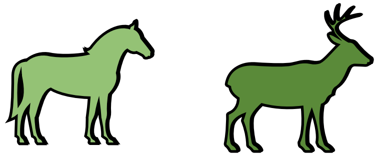
train the same
model also on
the new task



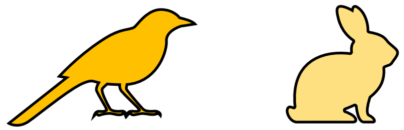
Task 1: "cat" or "dog"



Task 2: "horse" or "deer"



Task 3: "bird" or "rabbit"



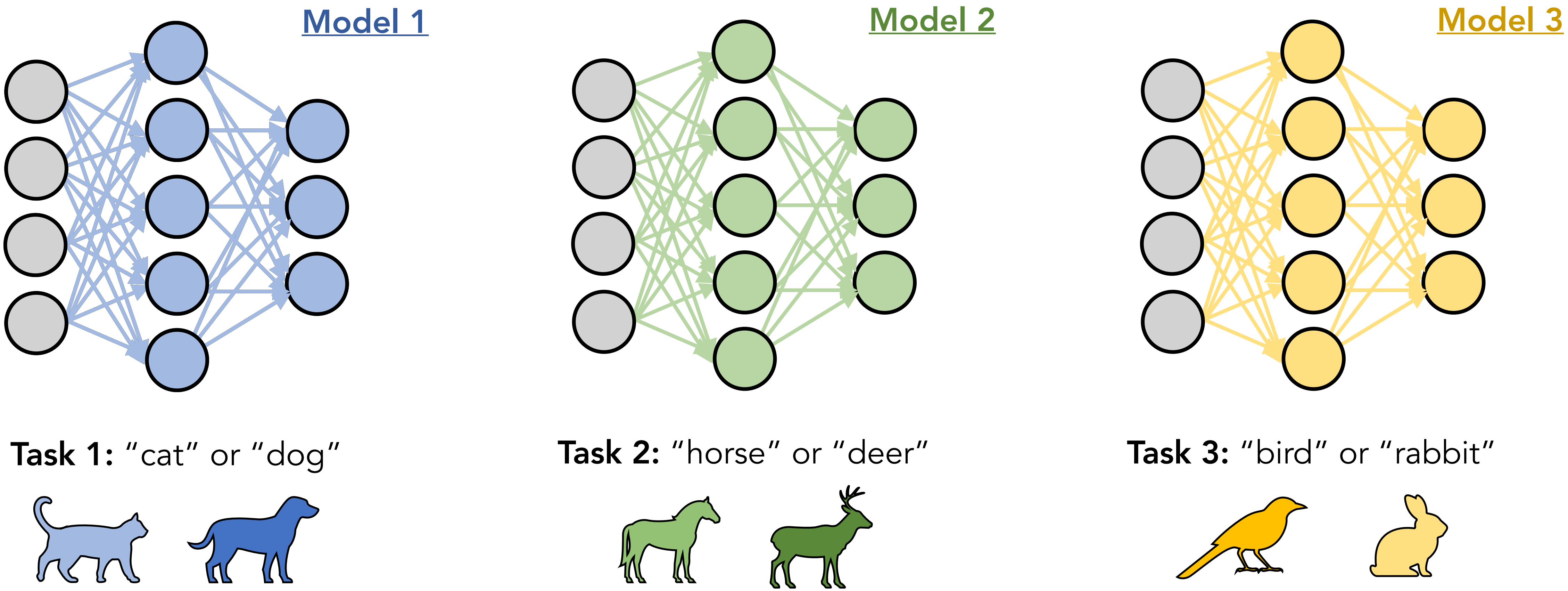
Main challenge: No access to previously observed datasets.

Commonly studied problem: Mitigating *catastrophic forgetting*.

Memory

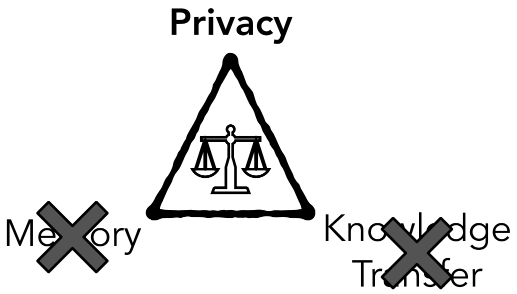
Knowledge
Transfer

Privacy-Aware Lifelong Learning

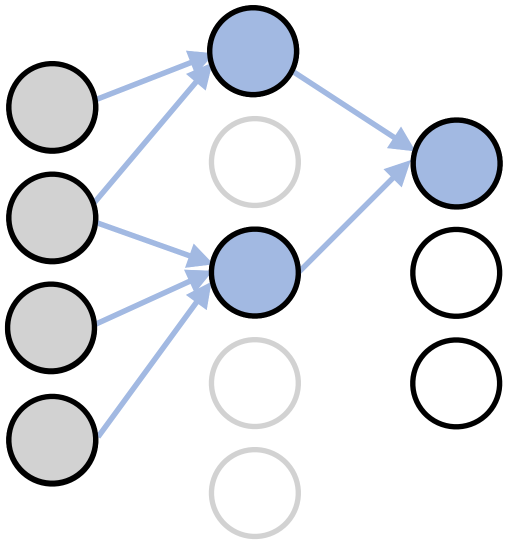


If we want to also ensure *exact unlearning* guarantees anytime:

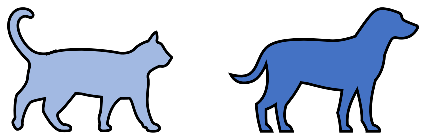
(a) Train different models for each task (brute force solution).



Privacy-Aware Lifelong Learning



Task 1: "cat" or "dog"



$$\min_{\boldsymbol{\theta},\mathbf{s}_t} \mathbb{E}_{(\boldsymbol{x},y)\sim\mathcal{D}^t}\left[\ell_{\mathrm{ce}}\left(\boldsymbol{x},y;\boldsymbol{\theta}\odot\mathbf{m}_t(\mathbf{s}_t)\right)\right]$$

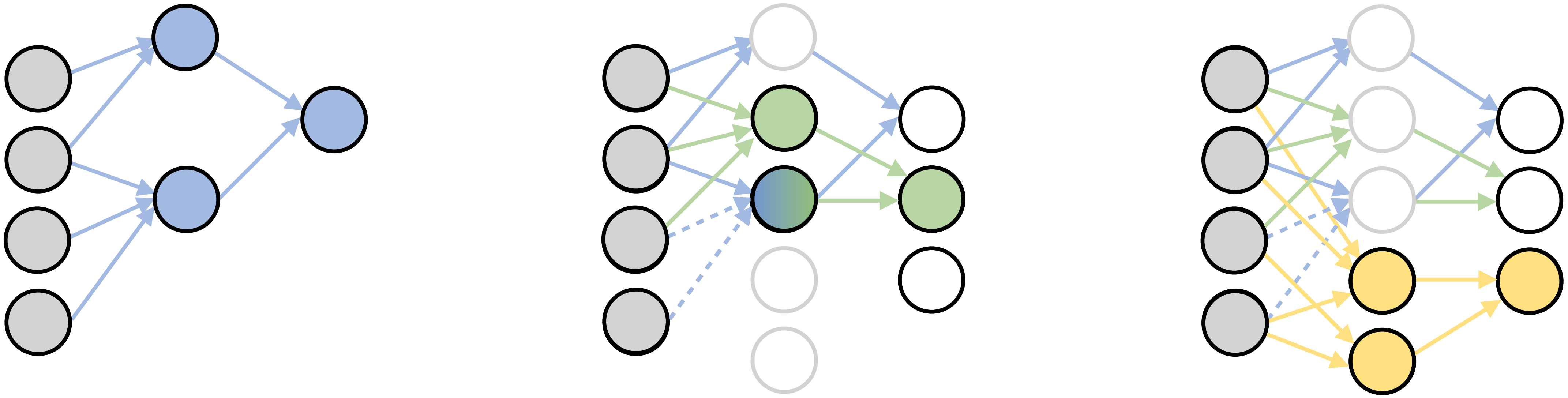
Enables **knowledge transfer**

$$\begin{cases} \mathbf{s}_t \leftarrow \mathbf{s}_t - \eta\left(\dfrac{\partial \ell_{\mathrm{ce}}}{\partial\, \mathbf{s}_t}\right) \\ \boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta\left(\dfrac{\partial \ell_{\mathrm{ce}}}{\partial\, \boldsymbol{\theta}} \odot (1 - \bigvee_j \mathbf{m}_j)\right) \end{cases}$$

Alleviates **catastrophic forgetting**

(b) Optimize *unlearnable* task-specific subnetworks! 👍

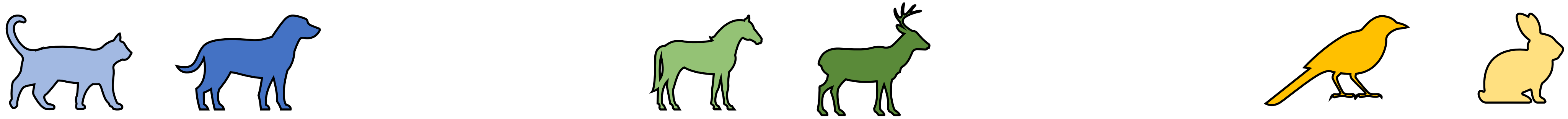
Privacy-Aware Lifelong Learning



Task 1: "cat" or "dog"

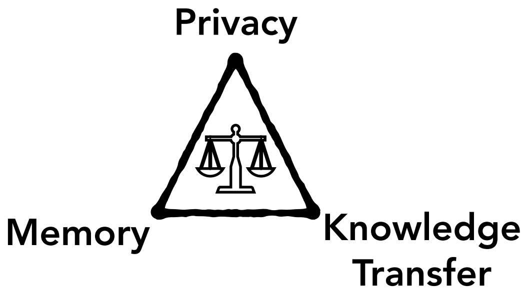
Task 2: "horse" or "deer"

Task 3: "bird" or "rabbit"

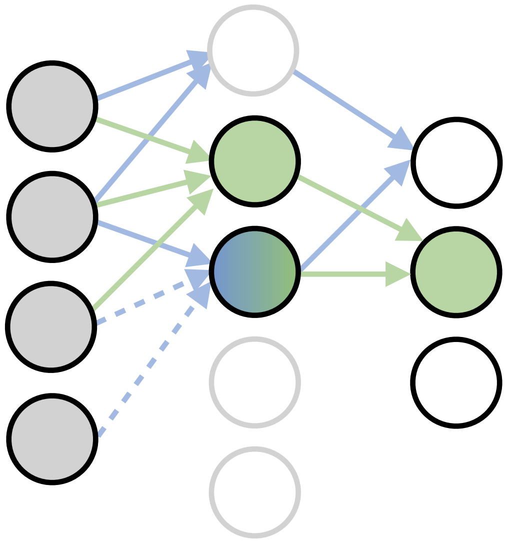


(b) Optimize *unlearnable* task-specific subnetworks!

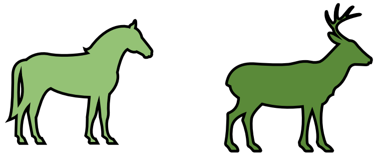
$$\min_{\boldsymbol{\theta}, \mathbf{s}_t} \mathbb{E}_{(\boldsymbol{x}, y) \sim \mathcal{D}^t} \left[\ell_{\mathrm{ce}} \left(\boldsymbol{x}, y; \boldsymbol{\theta} \odot \mathbf{m}_t(\mathbf{s}_t) \right) \right]$$

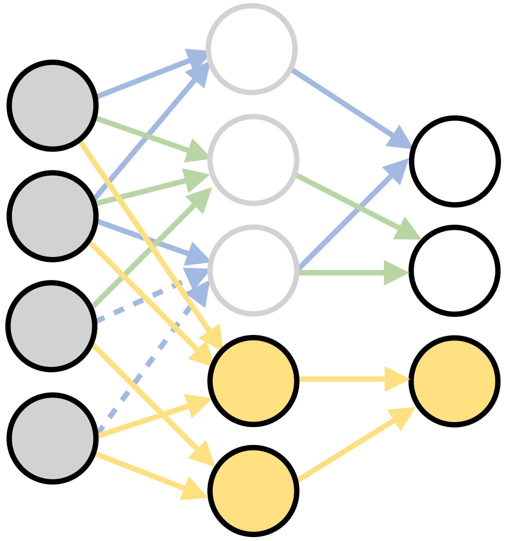


Privacy-Aware Lifelong Learning

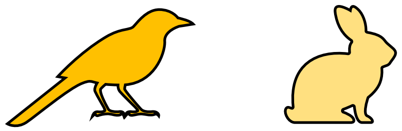


Task 2: "horse" or "deer"

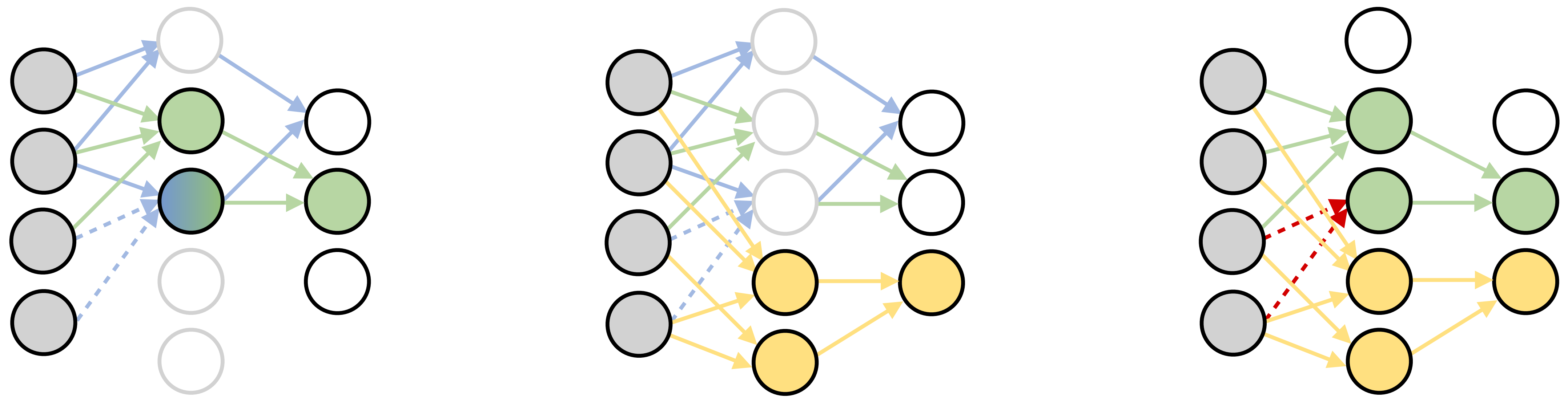




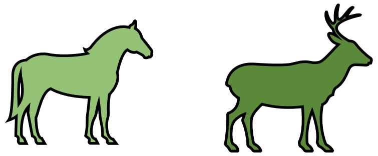
Task 3: "bird" or "rabbit"



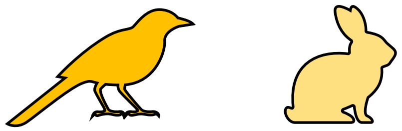
Privacy-Aware Lifelong Learning



Task 2: "horse" or "deer"



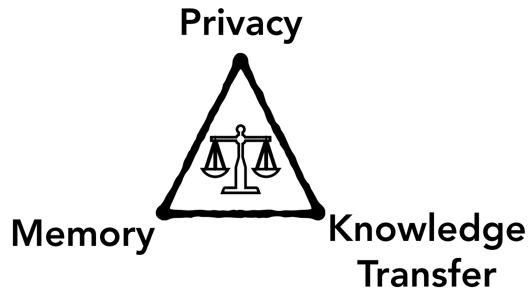
Task 3: "bird" or "rabbit"



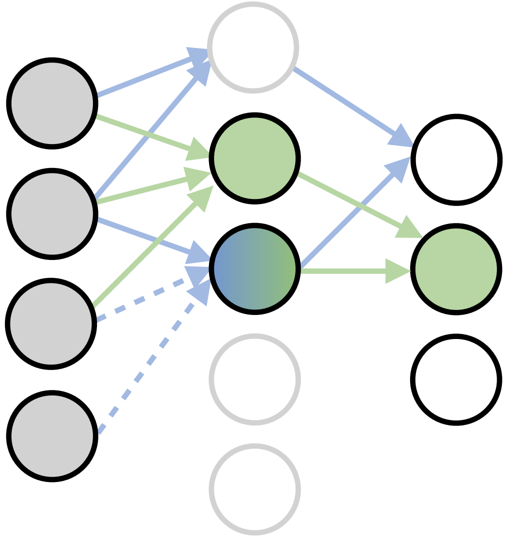
Unlearn Task 1
(Exact Unlearning)

- **Following a *task unlearning* request:**

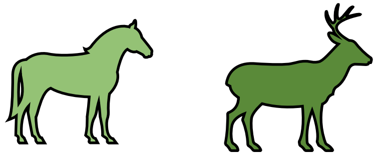
  Reset the corresponding subnetwork weights & use experience replay to retrain only **the weights** where knowledge was transferred from **Task 1**.

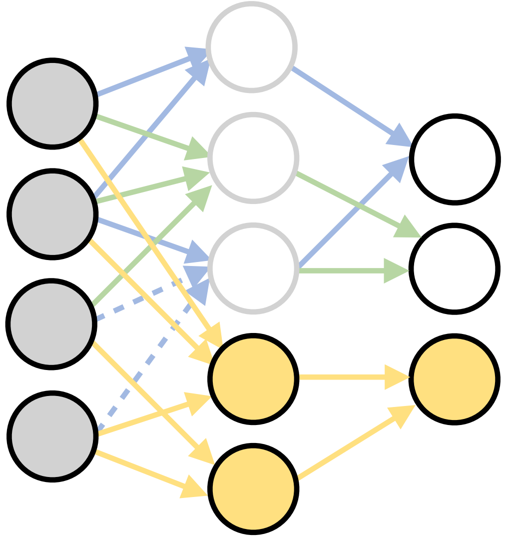


Privacy-Aware Lifelong Learning

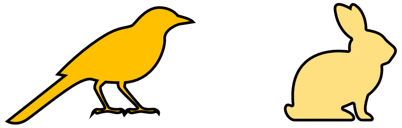


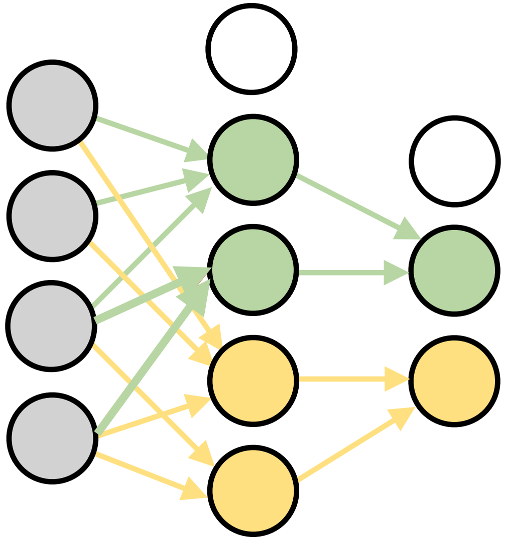
Task 2: "horse" or "deer"





Task 3: "bird" or "rabbit"

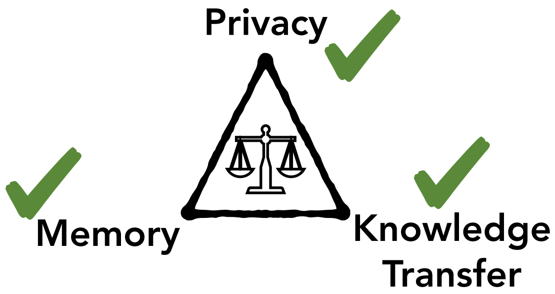




Unlearn Task 1
(Exact Unlearning)

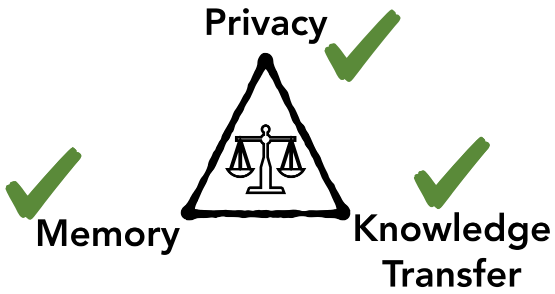
- **Following a *task unlearning* request:**

Reset the corresponding subnetwork weights & use experience replay to retrain only **the weights** where knowledge was transferred from **Task 1**.



Summary & Takeaways

- Resource-efficiency in deep learning is important (e.g., **sustainable AI, embedded DL**, …).
- Dedicated algorithms are necessary for ***resource-efficient*** and ***reliable deep learning***.
- ***Privacy-aware lifelong learning (PALL)*** is designed to:
  - **Alleviate catastrophic forgetting** by freezing pre-trained task-specific weights,
  - Facilitate **selective knowledge transfer** from previously learned tasks,
  - **Ensure exact task unlearning** guarantees upon request,
  - Provide a state-of-the-art solution with **minimal model memory overhead**.



Thank you for your attention!

MONTAN UNIVERSITÄT LEOBEN

TU Graz

FWF Der Wissenschaftsfonds.

GraML Graz Center for Machine Learning

Acknowledgments: This research was funded in whole or in part by the **Austrian Science Fund (FWF)** [10.55776/COE12]. This work has also been supported by the **Graz Center for Machine Learning (GraML)**.

PRIVACY-AWARE LIFELONG LEARNING

Ozan Özdenizci[1], **Elmar Rueckert**[1], **Robert Legenstein**[2]
[1] Chair of Cyber-Physical-Systems, Montanuniversität Leoben, Austria
[2] Institute of Machine Learning and Neural Computation, Graz University of Technology, Austria
{ozan.oezdenizci,elmar.rueckert}@unileoben.ac.at
robert.legenstein@tugraz.at

ABSTRACT

Lifelong learning algorithms enable models to incrementally acquire new knowledge without forgetting previously learned information. Contrarily, the field of machine unlearning focuses on explicitly forgetting certain previous knowledge from pretrained models when requested, in order to comply with data privacy regulations on the *right-to-be-forgotten*. Enabling efficient lifelong learning with the capability to selectively unlearn sensitive information from models presents a critical and largely unaddressed challenge with contradicting objectives. We address this problem from the perspective of simultaneously *preventing catastrophic forgetting* and *allowing forward knowledge transfer* during task-incremental learning, while *ensuring exact task unlearning* and *minimizing memory requirements*, based on a single neural network model to be adapted. Our proposed solution, privacy-aware lifelong learning (PALL), involves optimization of task-specific sparse subnetworks with parameter sharing within a single architecture. We additionally utilize an episodic memory rehearsal mechanism to facilitate exact unlearning without performance degradations. We empirically demonstrate the scalability of PALL across various architectures in image classification, and provide a state-of-the-art solution that uniquely integrates lifelong learning and privacy-aware unlearning mechanisms for responsible AI applications.

1 INTRODUCTION

Lifelong learning algorithms enhance the ability of machine learning models to incrementally acquire new skills or integrate new knowledge over time from sequentially observed data (van de Ven et al., 2022). This continual learning capability is essential for models to stay relevant in dynamic environments where the observed data distributions change. A widely studied challenge in this setting is to mitigate *catastrophic forgetting*, addressing the loss of prior knowledge as new tasks are learned. There has been various strategies proposed to prevent forgetting, while exploiting *forward knowledge transfer* to efficiently improve performance in new tasks. However, these lifelong learning approaches conventionally do not consider the factor of ensuring data privacy, whereas selectively forgetting (or *unlearning*) certain knowledge may be required to comply with the legal regulations on the *right-to-be-forgotten* (Mantelero, 2013) (e.g., deleting prior information from personalized recommendation systems). This introduces an additional dimension of complexity, which requires novel lifelong learning solutions that can ensure unlearning for privacy-awareness.

The field of machine unlearning focuses on explicitly removing the influence of specific data points from pretrained models (Cao & Yang, 2015). Ensuring *exact unlearning*, where the model is guaranteed to behave as if the unlearned data was never observed, presents a significant challenge that generally requires partial model retraining (Bourtoule et al., 2021). In particular, current unlearning solutions assume previous or all data to be available to facilitate exact unlearning, which does not apply to lifelong learning settings where the data is only sequentially observed. Accordingly, recent works have started to explore solutions at the intersection of task-incremental lifelong learning and machine unlearning (Shibata et al., 2021; Liu et al., 2022; Chatterjee et al., 2024), primarily via inexact unlearning methods which does not guarantee privacy for all previously learned tasks.

We consider a similar lifelong learning problem, where the learning sequence may include *exact* task unlearning requests for any of the previously learned tasks, with no access to prior data. A