Open World Lifelong Learning
A Continual Machine Learning Course

Teacher

Dr. Martin Mundt,

hessian.AI-DEPTH junior research group leader on Open World Lifelong Learning (OWLL)

& researcher in the Artificial Intelligence and Machine Learning (AIML) group at TU Darmstadt

Time

Every Tuesday 17:30 - 19:00 CEST

Course Homepage

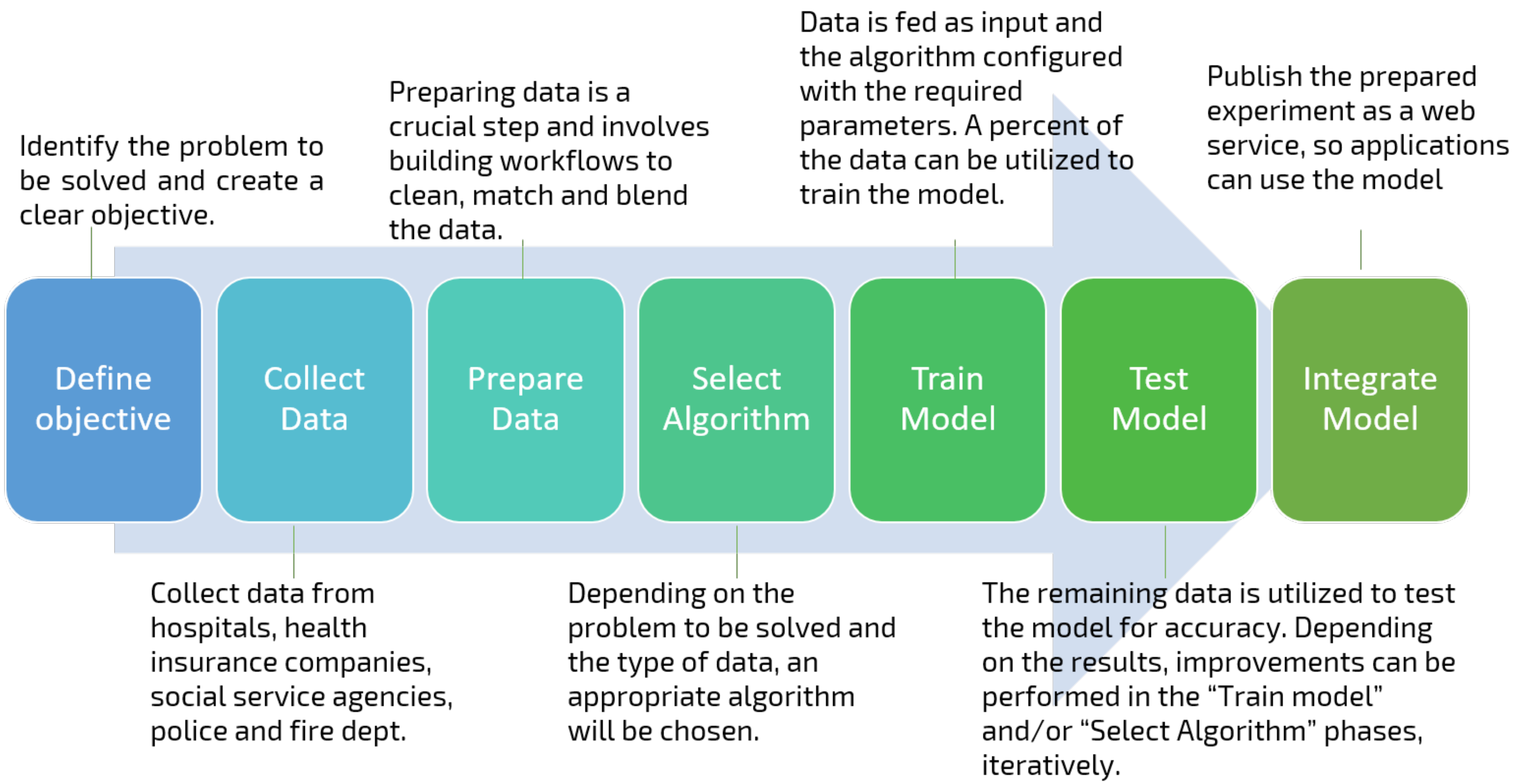
http://owll-lab.com/teaching/cl_lecture

https://www.youtube.com/playlist?list=PLm6QXeaB-XkA5-IVBB-h7XeYzFzgSh6sk

Week 12: Course wrap-up & Frontiers
+ Q&A session at the end

We have started with the question
What do you think: what is machine learning?

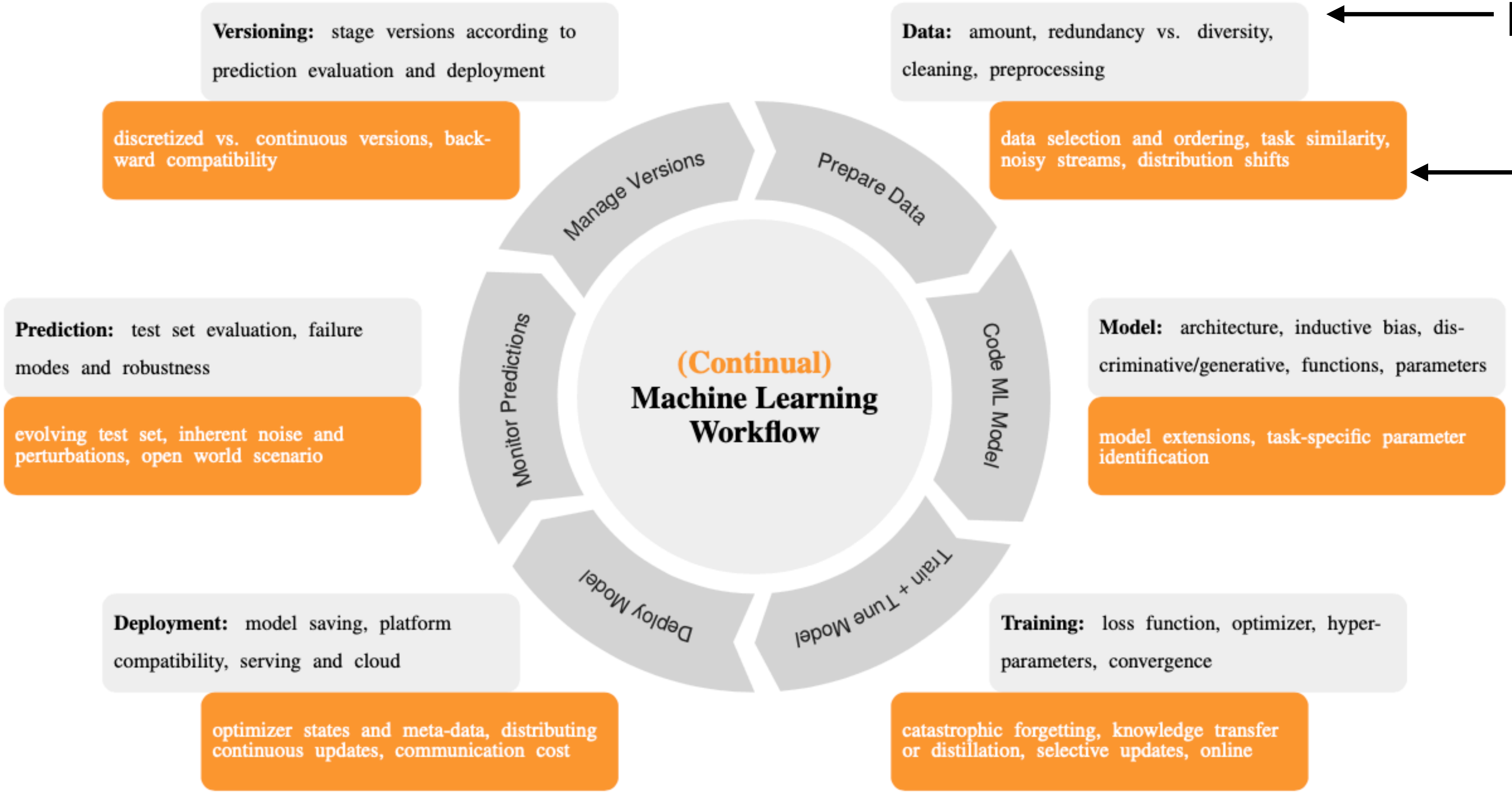
Can we just iterate?



We've quickly learned that it's more than "train-val-test"

Figure from https://www.congrelate.com/get-workflow-machine-learning-images/

Perhaps harder than expected?

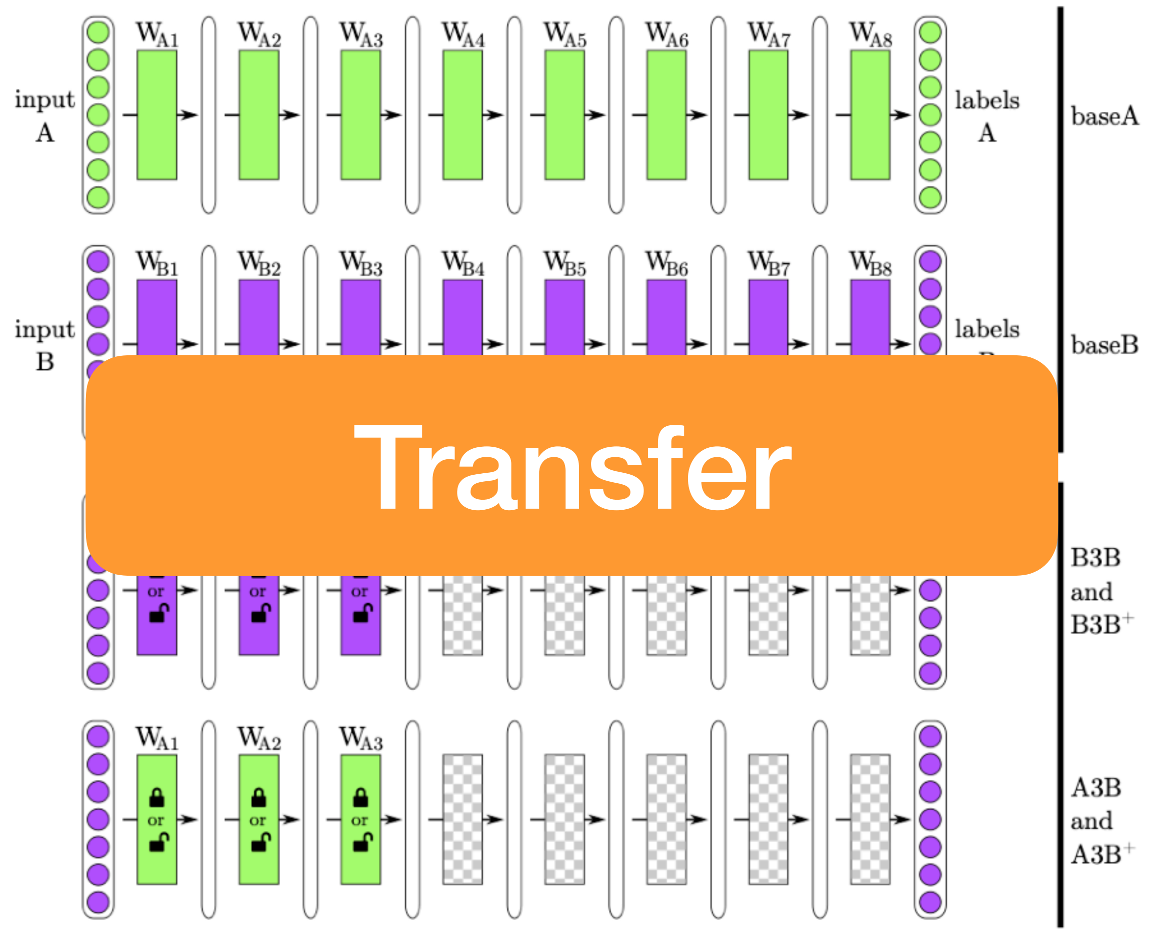


Individual questions

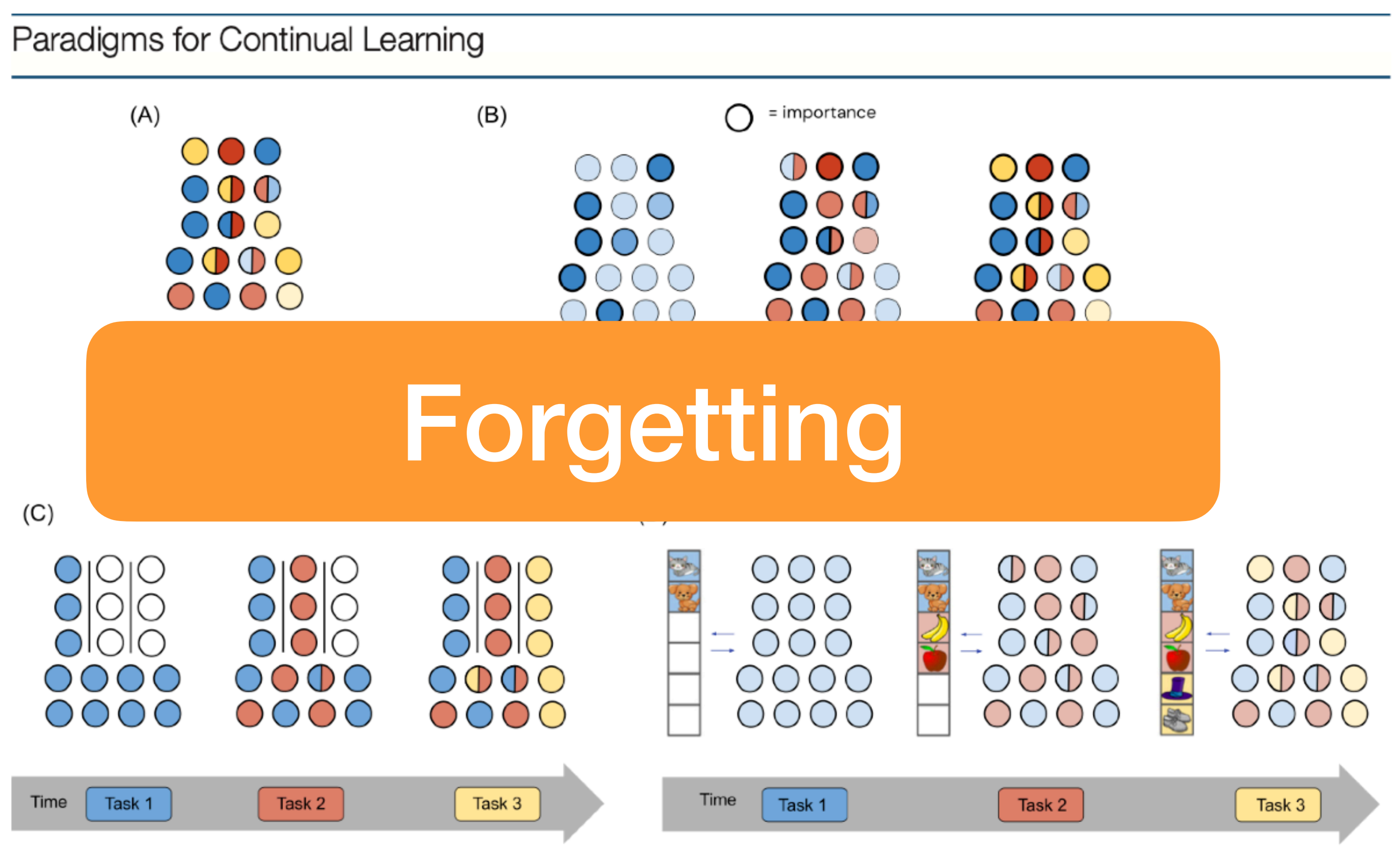
Continual dependencies & synergies

Mundt et al, “CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability”, ICLR 2022

What we've talked about



"How transferable are features in deep neural networks",
Yosinski et al, NeurIPS 2014



Hadsell et al, "Embracing Change: Continual Learning in Deep Neural Networks", Trends in Cognitive Sciences 24:12, 2020

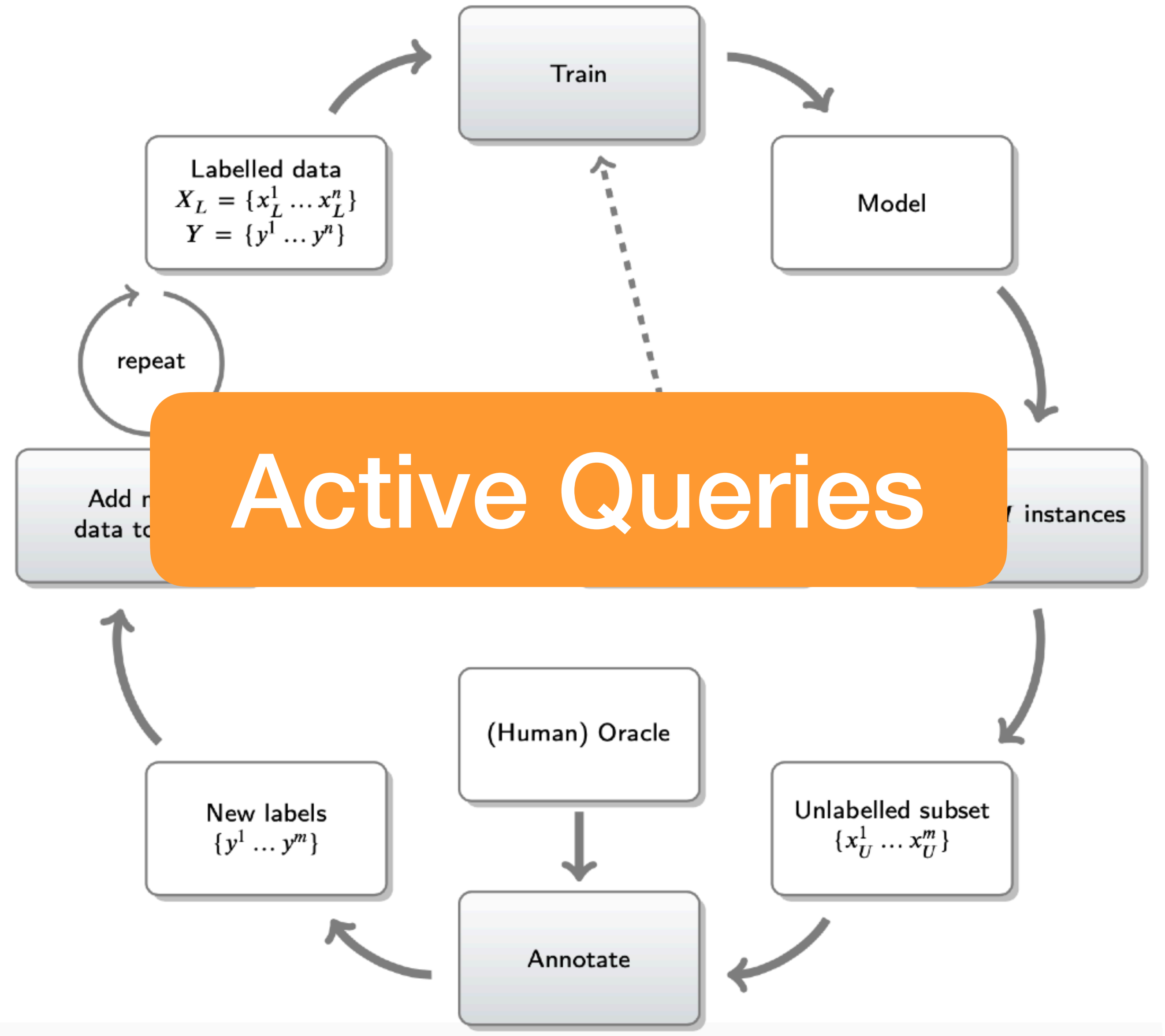
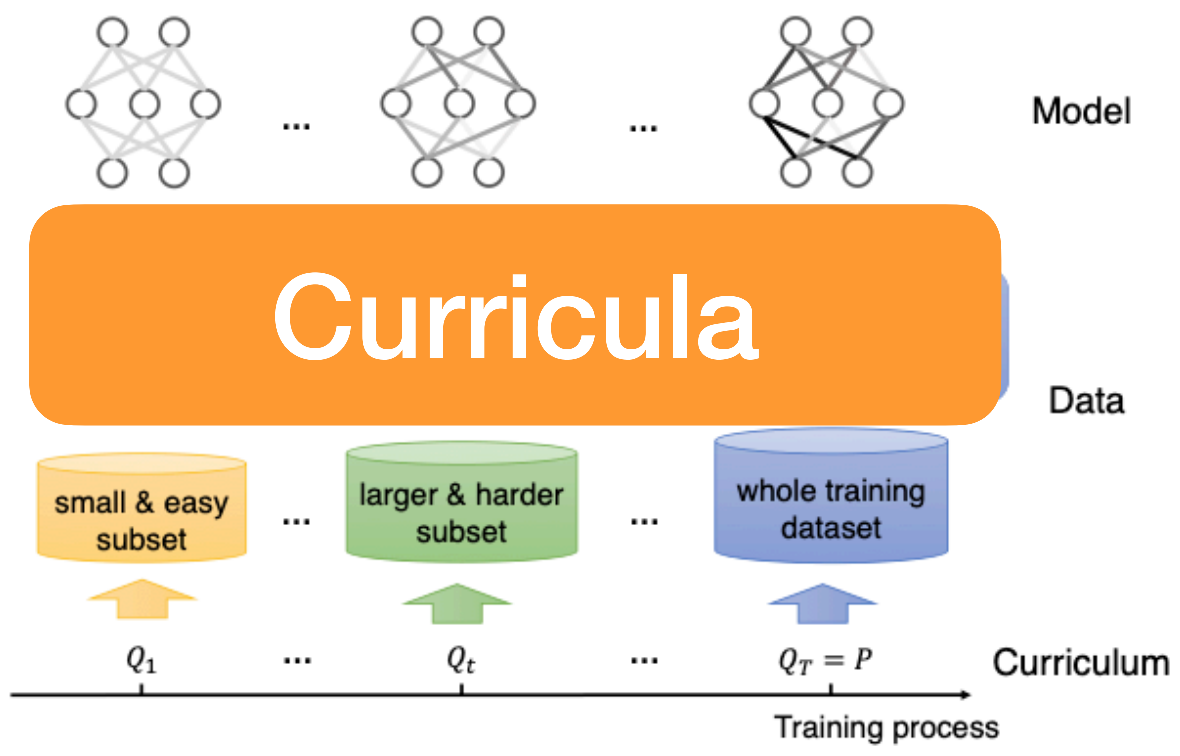


Figure from "A Wholistic View of Deep Neural Networks: Forgotten Lessons and the Bridge to Active and Open World Learning", Mundt et al 2020



Wang et al, "A Survey on Curriculum Learning", TPAMI 2021

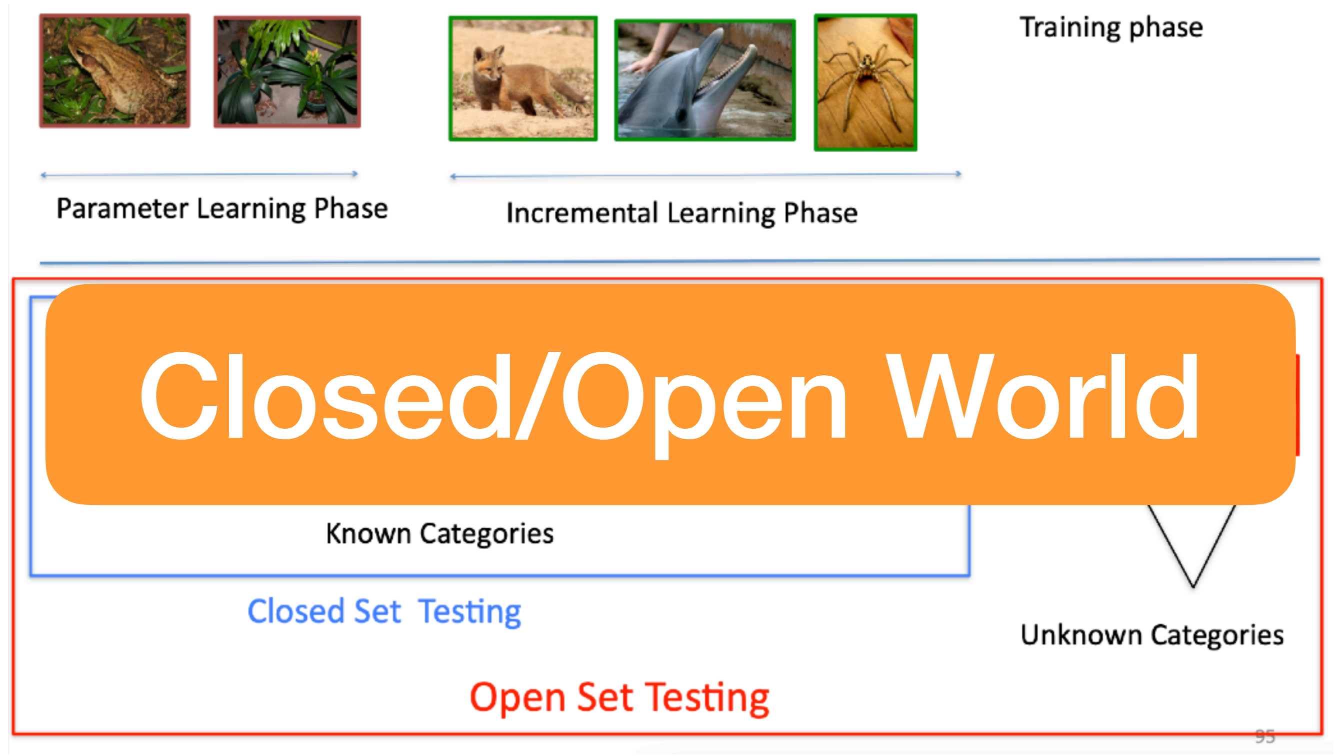
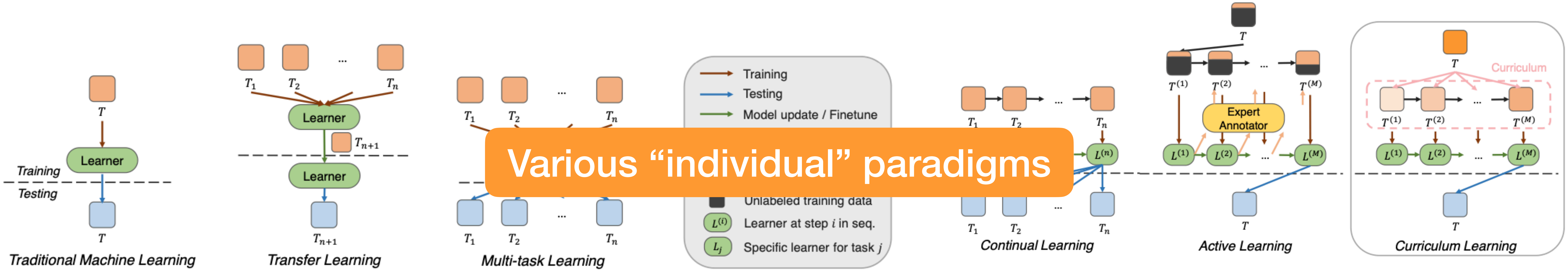


Figure from CVPR16 "Statistical Methods for Open Set Recognition" by Scheirer & Boult, https://www.wjscheirer.com/misc/openset/cvpr2016-open-set-part3.pdf

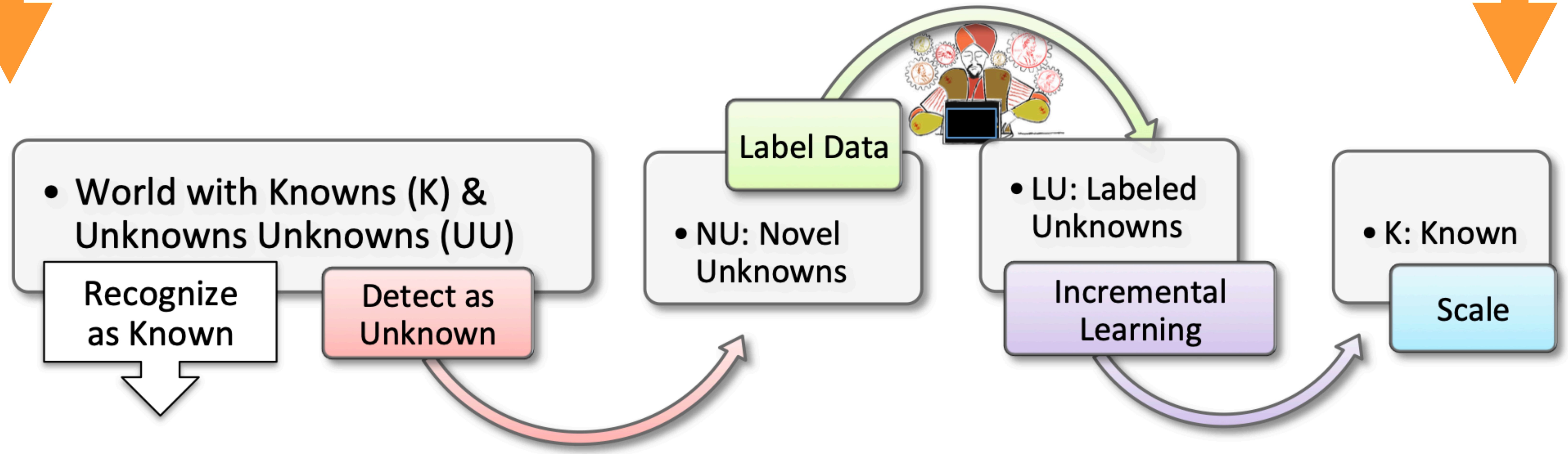
What we've talked about



Various "individual" paradigms

Wang et al, "A Survey on Curriculum Learning", TPAMI 2021

Are puzzle pieces to a "lifelong open world learner"

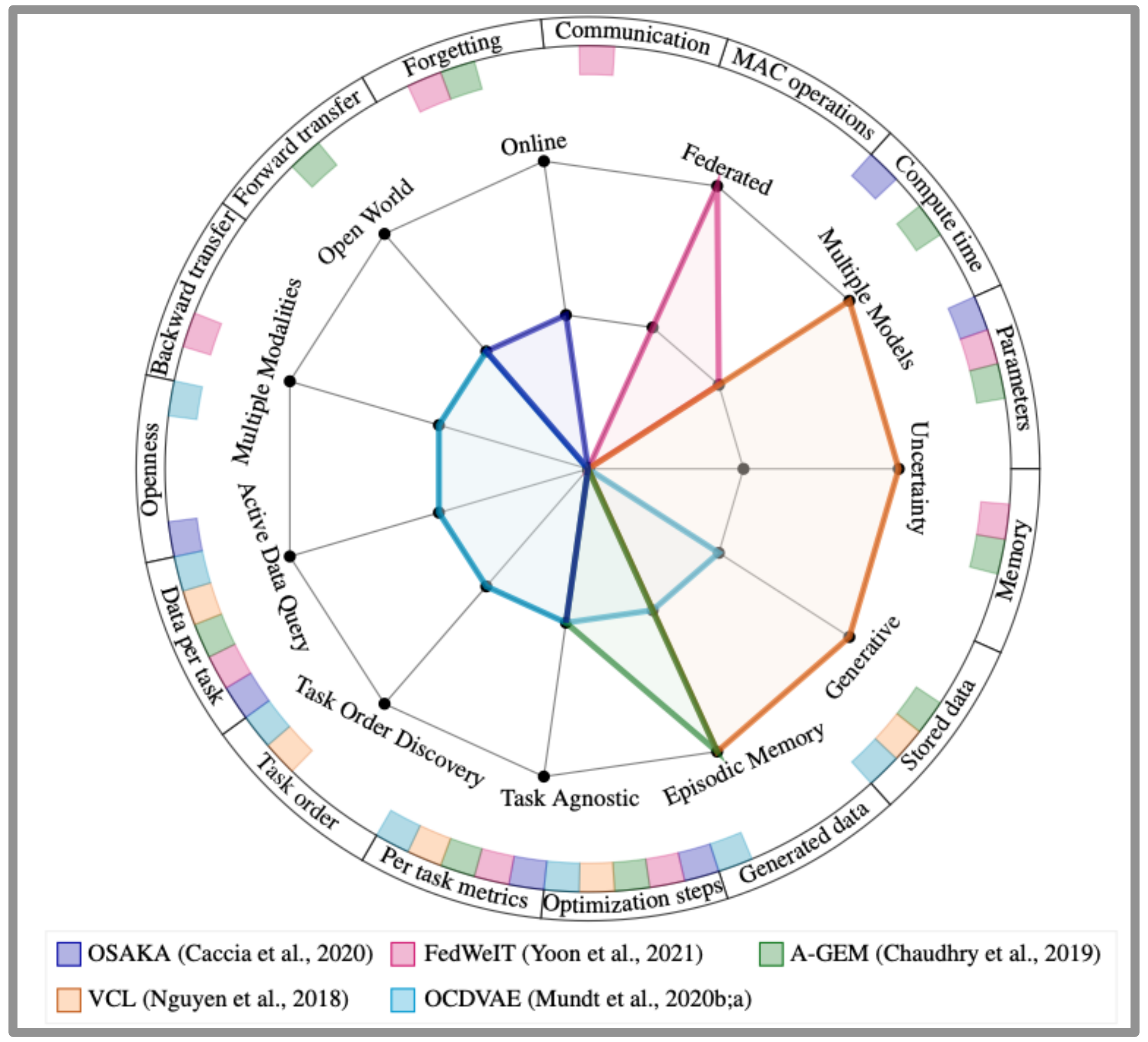


Bendale & Boult ,"Towards Open World Recognition", CVPR 2015

What we’ve talked about

OWLL ContinualAI hessian.AI TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluation



Mundt et al, “CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability”, ICLR 2022

Soft+Hardware

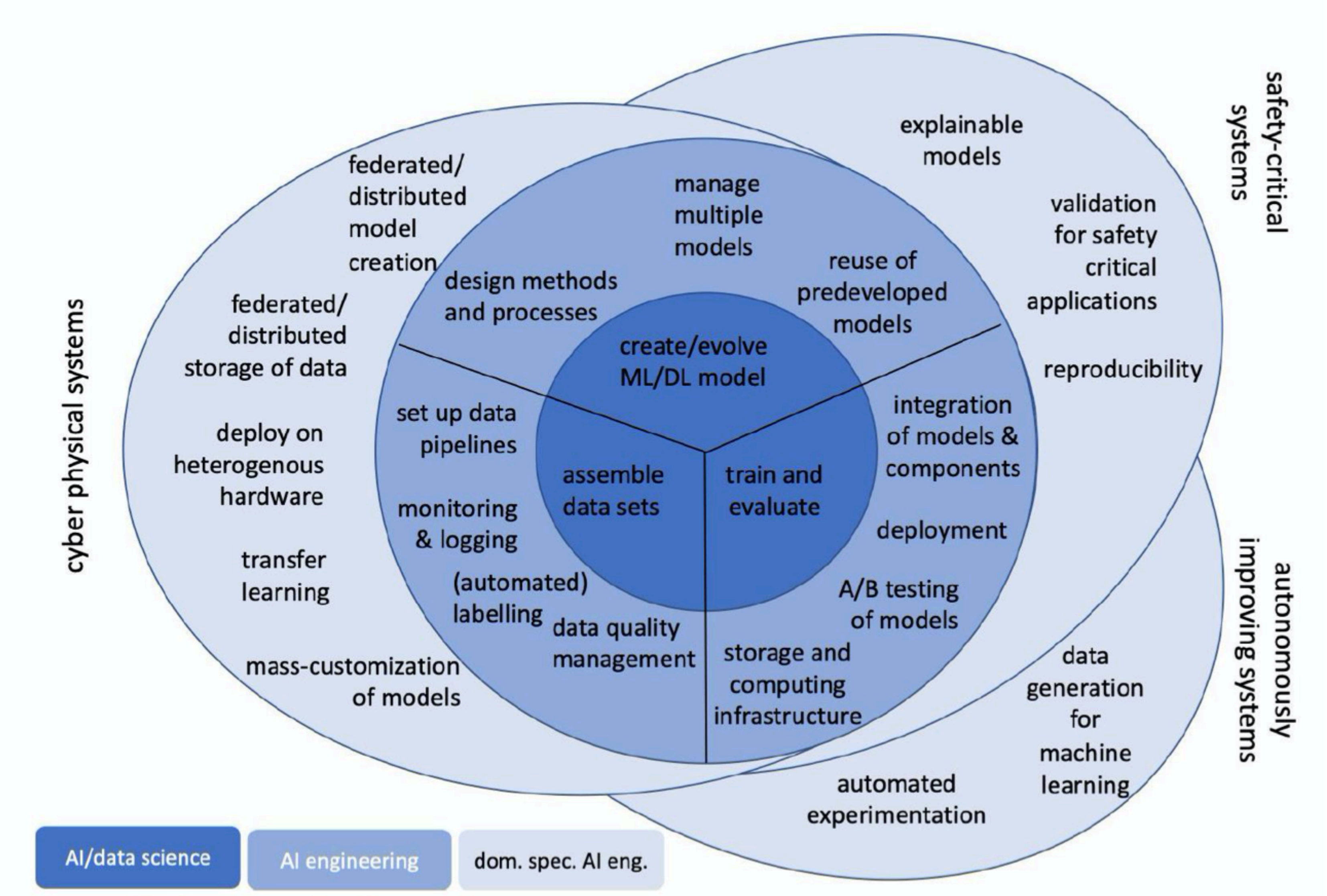
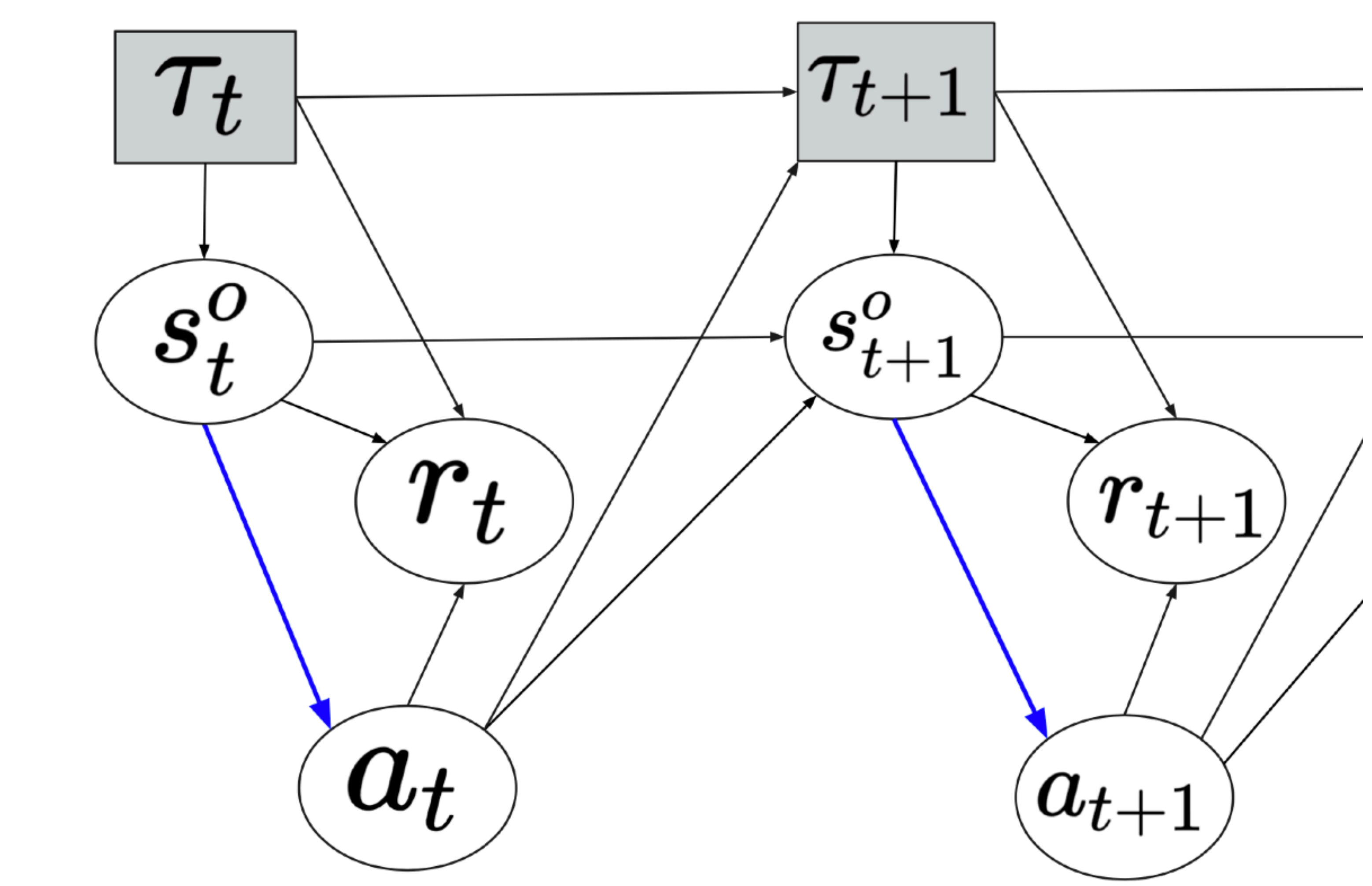


Figure 3: Conceptualization of AI engineering

Bosch et al, “Engineering AI Systems: A Research Agenda”, in Artificial Intelligence Paradigms for Smart Cyber-Physical Systems

Supervision/Reinforcement

TACRL as a POMDP special case



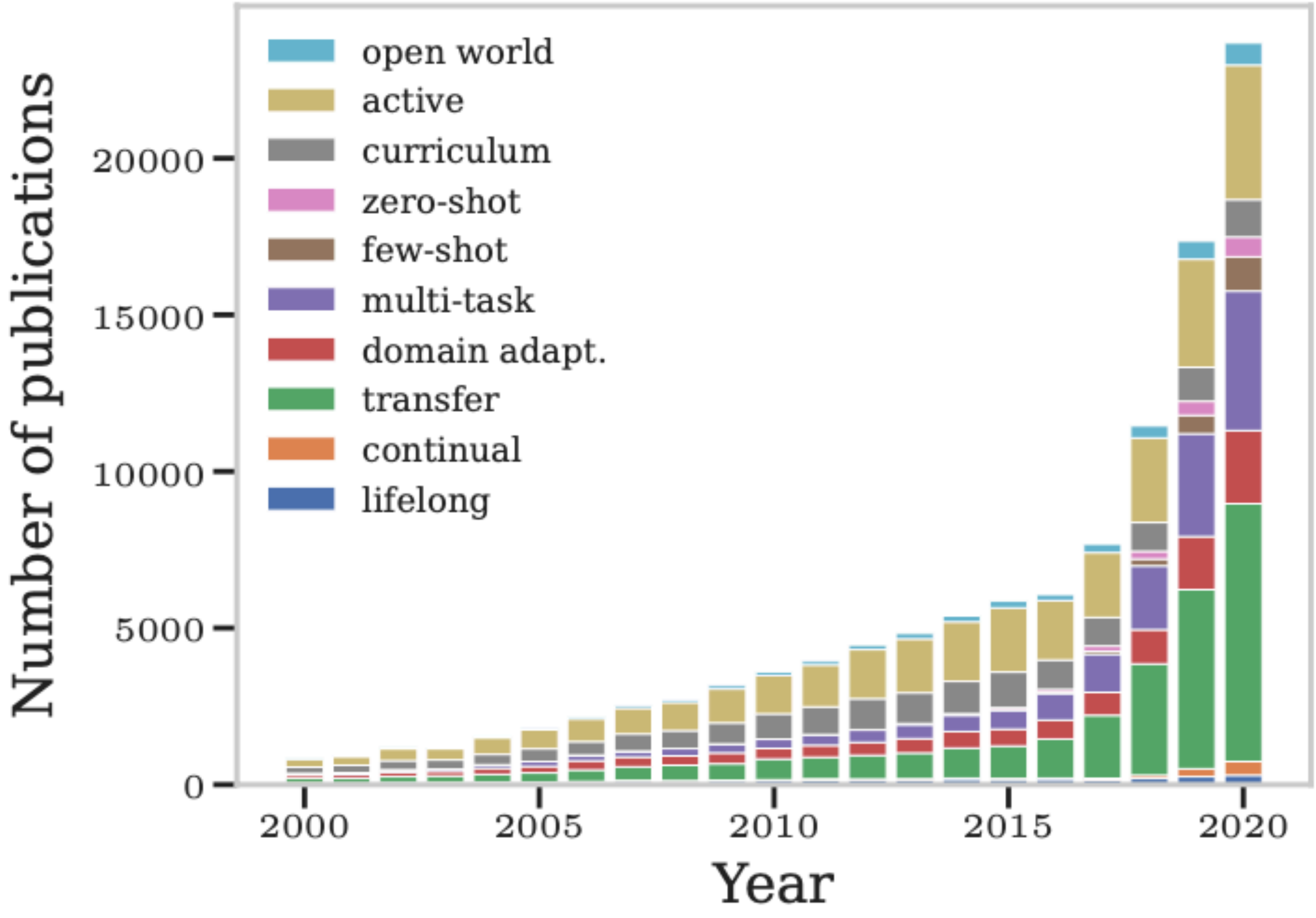
From guest lecture, week 11, task-agnostic reinforcement learning

We’ve already encountered many frontiers

Each “individual paradigm” has its frontiers,
even before drawing connections

A central question seems to be a trade-off?
The value of the “whole” & the utility of a “niche”

Dependencies & synergies



We should now be more familiar with the left picture

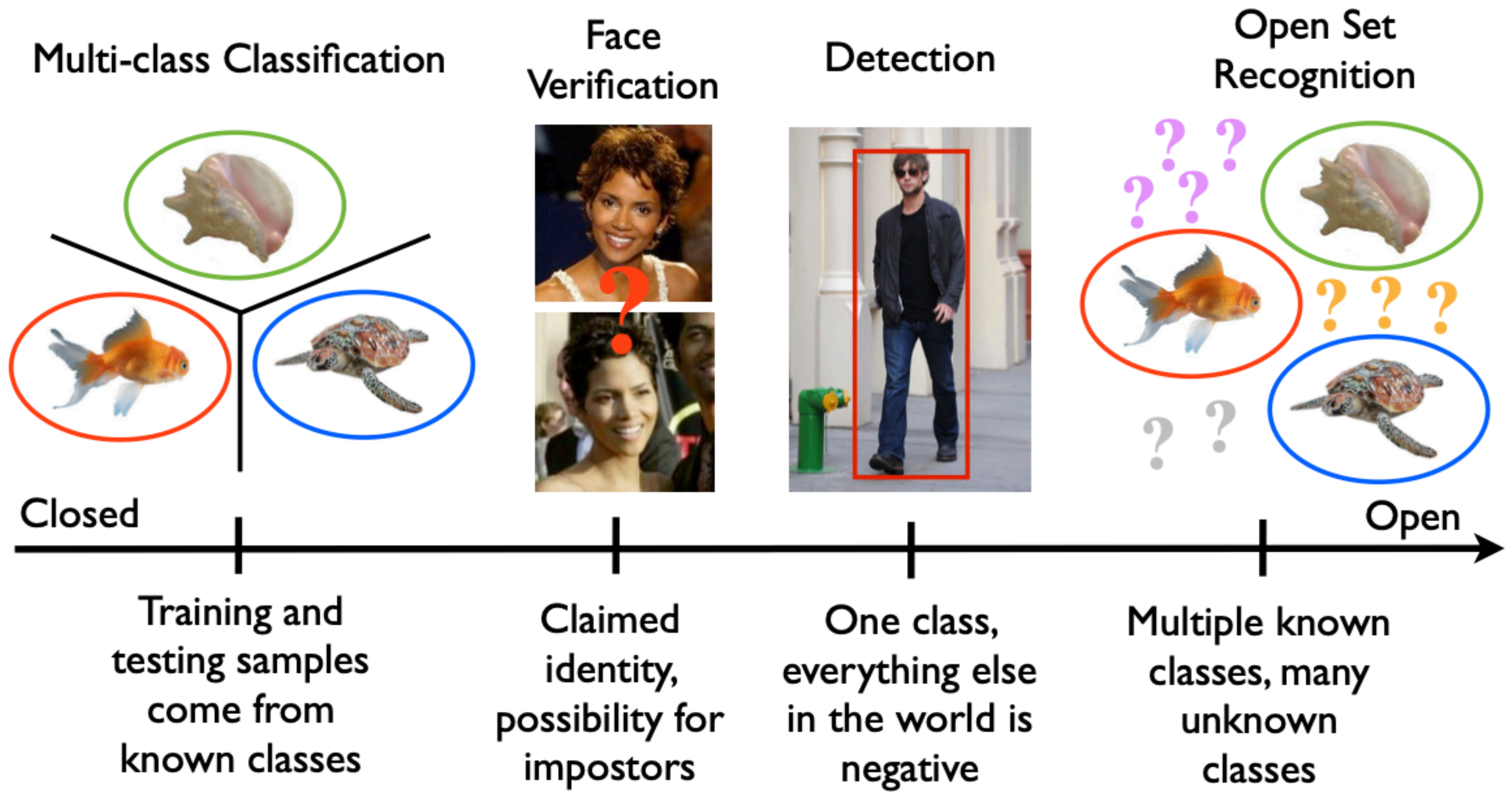
And hopefully also have some understanding of the dependencies, the complex interplay & existing synergies

Mundt et al, "CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability", ICLR 2022

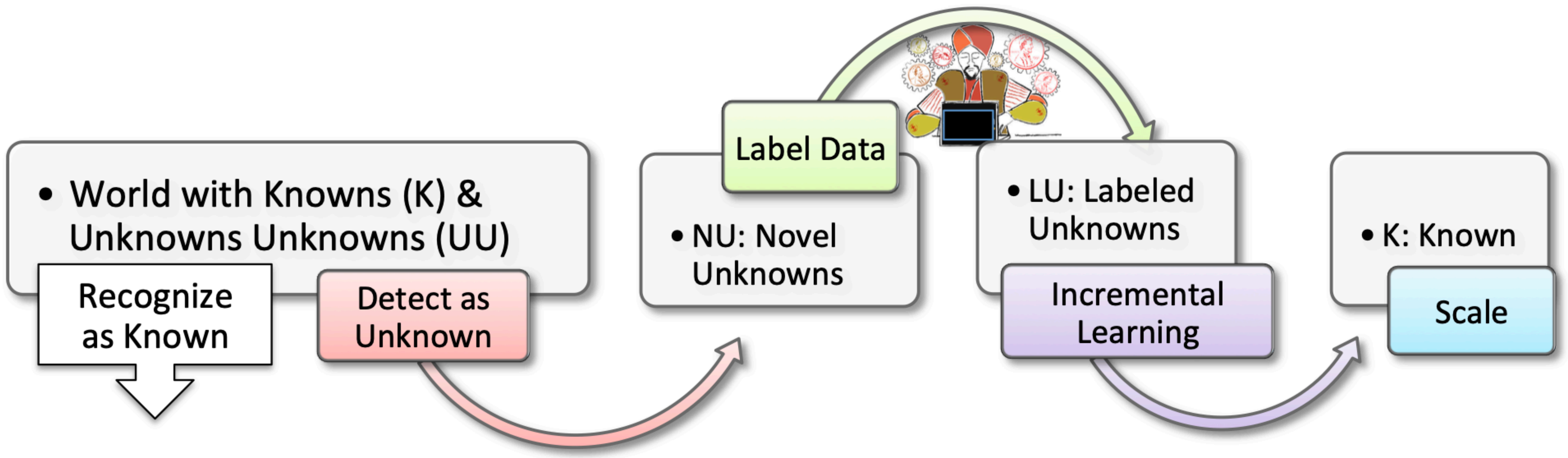
Closed vs. open worlds

It's likely we will need to study both: specifics + overall systems!

But when do we study what? And when are our assumptions fair?



Scheirer et al, "Towards Open Set Recognition", TPAMI 2012



Bendale & Boult ,"Towards Open World Recognition", CVPR 2015

Evaluation & related paradigms

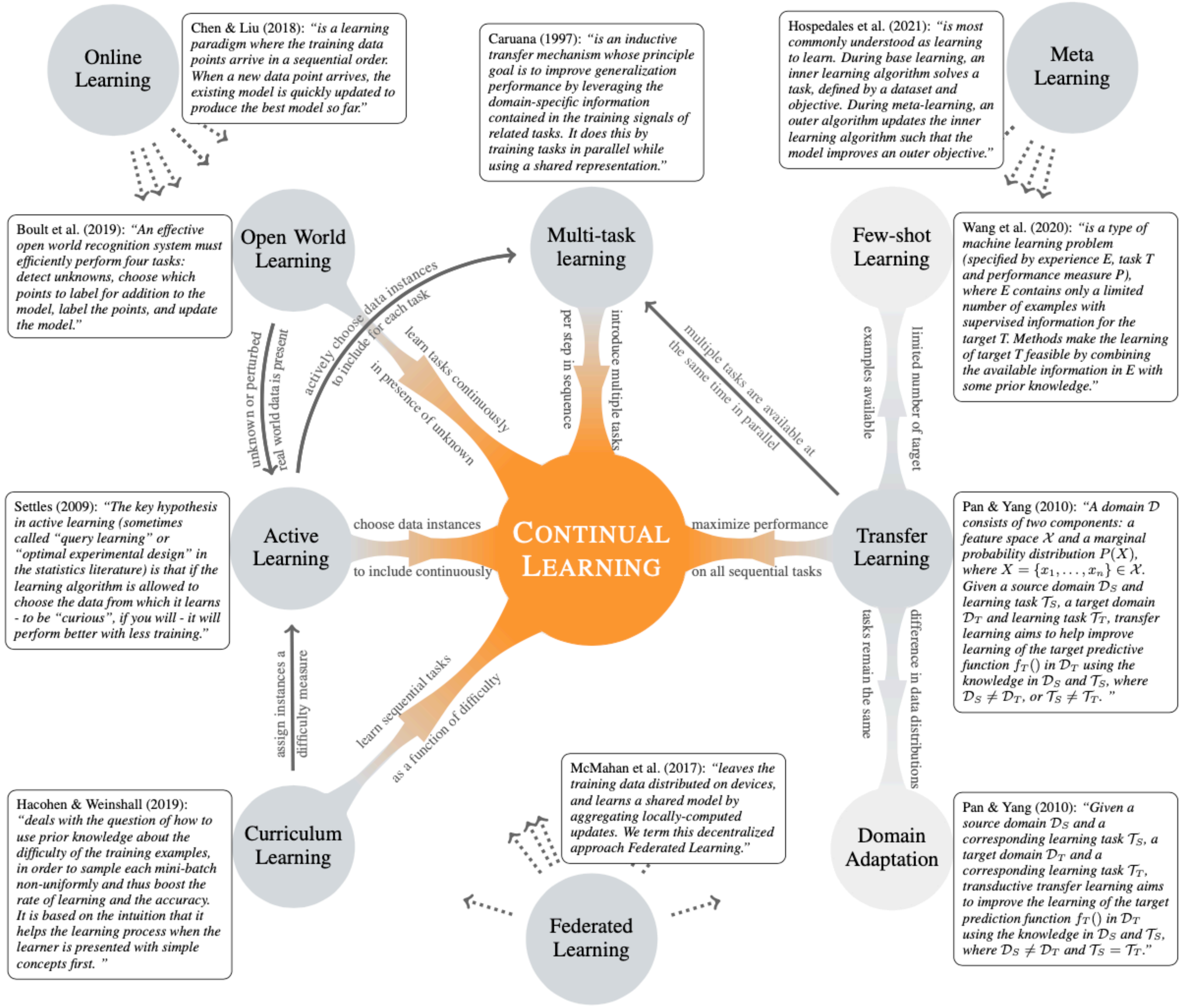
The differences between machine learning paradigms with continuous components can be nuances

Key aspects often reside in how we evaluate

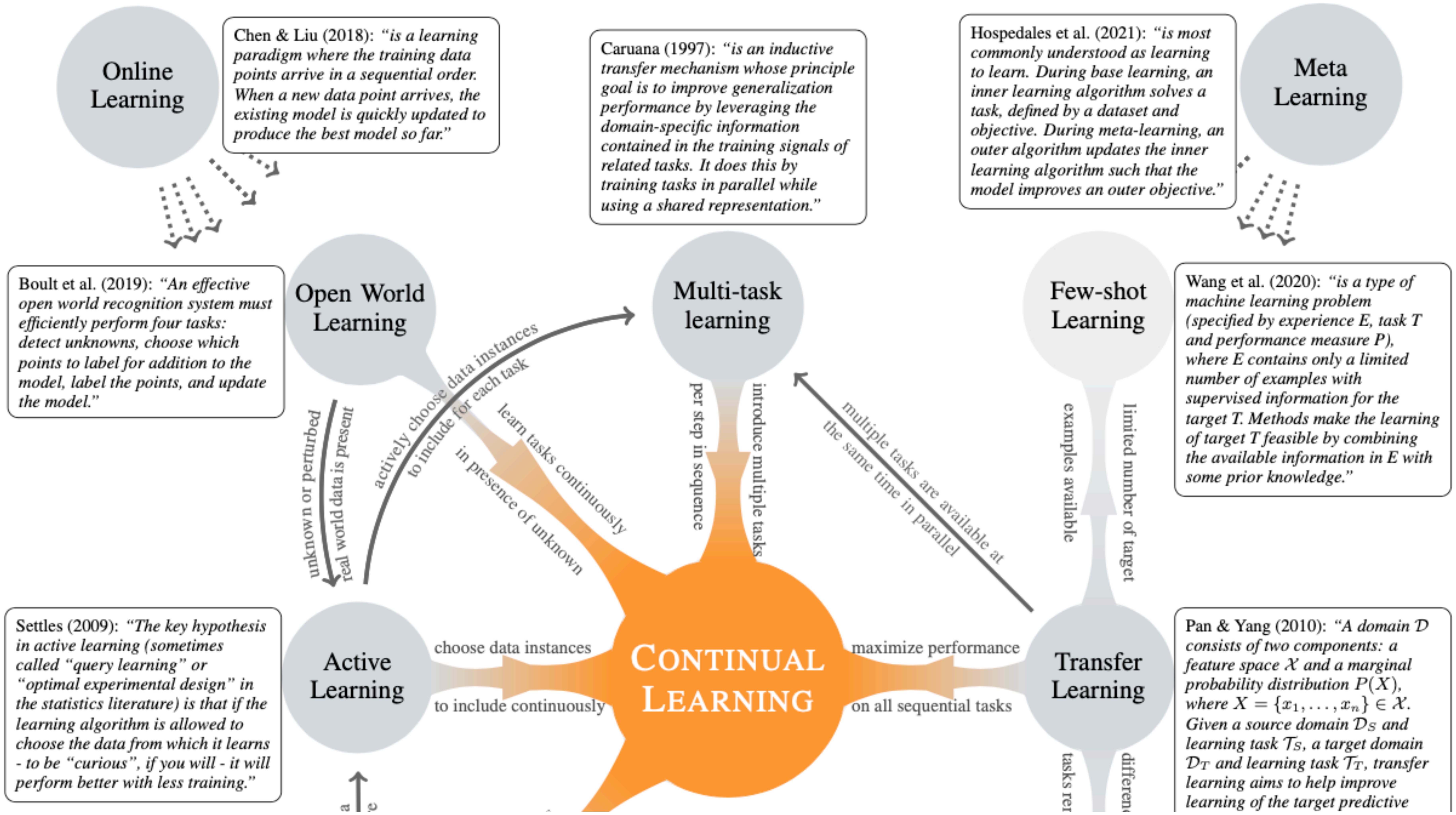
Each paradigm seems to have a particular preference (potentially neglecting other important factors)

Assumptions, benchmarks & evaluation in themselves are a frontier!

Mundt et al, "CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability", ICLR 2022

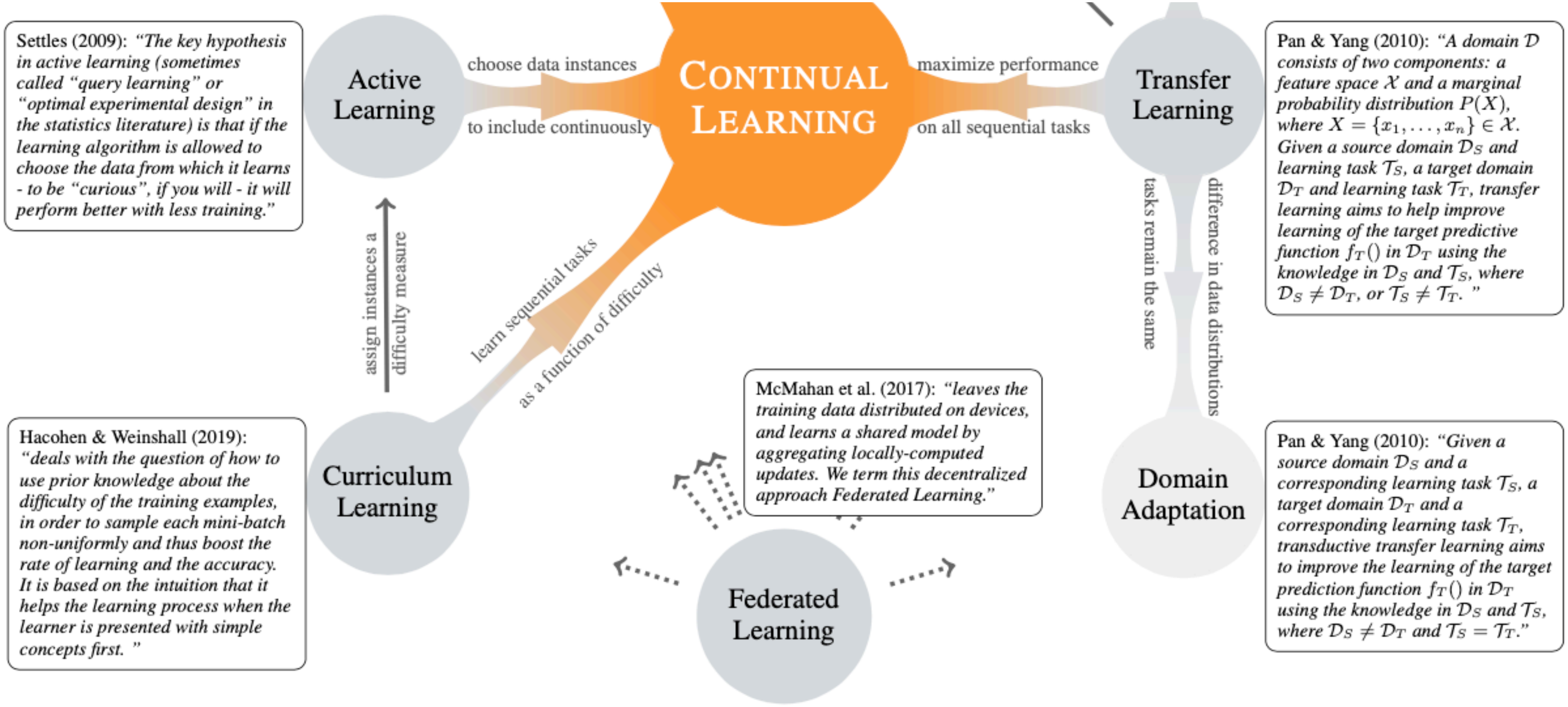


Evaluation & related paradigms



Mundt et al, "CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability", ICLR 2022

Evaluation & related paradigms



Mundt et al, "CLEVA-Compass: A Continual Learning Evaluation Assessment Compass to Promote Research Transparency and Comparability", ICLR 2022

Early definition: lifelong ML

Provocatively asking:

Is it even possible/desirable to strive for a unified definition of lifelong machine learning?

Definition - Lifelong Machine Learning - Thrun 1996:

"The system has performed N tasks. When faced with the (N+1)th task, it uses the knowledge gained from the N tasks to help the (N+1)th task."

"Is Learning The n-th Thing Any Easier Than Learning the First?" (NeurIPS 1996) & "Explanation based Neural Network Learning A Lifelong Learning Approach", Springer US, 1996

Later definition: lifelong ML

Definition - **Lifelong Machine Learning** - Chen & Liu 2017:

"Lifelong Machine Learning is a continuous learning process. At any time point, the learner performed a sequence of N learning tasks, $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N$ *.These tasks can be of the same type or different types and from the same domain or different domains. When faced with the (N+1)th task* $\mathcal{T}_{N+1}$ *(which is called the new or current task) with its data* D_{N+1}*, the learner can leverage past knowledge in the knowledge base (KB) to help learn* $\mathcal{T}_{N+1}$*.*

The objective of LML is usually to optimize the performance on the new task $\mathcal{T}_{N+1}$*, but it can optimize any task by treating the rest of the tasks as previous tasks. KB maintains the knowledge learned and accumulated from learning the previous task. After the completion of learning* $\mathcal{T}_{N+1}$*, KB is updated with the knowledge (e.g. intermediate as well as the final results) gained from learning* $\mathcal{T}_{N+1}$*. The updating can involve inconsistency checking, reasoning, and meta-mining of additional higher-level knowledge."*

"Lifelong Machine Learning", Chen & Liu, Morgan Claypool, 2017

Later definition: lifelong ML

Definition - Lifelong Machine Learning - Chen & Liu 2017:

"Lifelong Machine Learning is a continuous learning process. At any time point, the learner performed a sequence of N learning tasks, $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N$. These tasks can be of the same

type o … ced with

the (N … $_{N+1}$, the

learne … $_{+1}$.

The o … $_{+1}$, but it

can op … tains the

knowle … mpletion

of lear … the final

results) gained from learning $\mathcal{T}_{N+1}$. The updating can involve inconsistency checking, reasoning, and meta-mining of additional higher-level knowledge."

May contain some parts we haven't discussed: reasoning, meta-mining of higher-level knowledge …

Does not explicitly contain many things we have learned about: active data queries, difficulty/curricula, dynamic model architectures, open worlds, soft/hardware, memory/compute constraints …

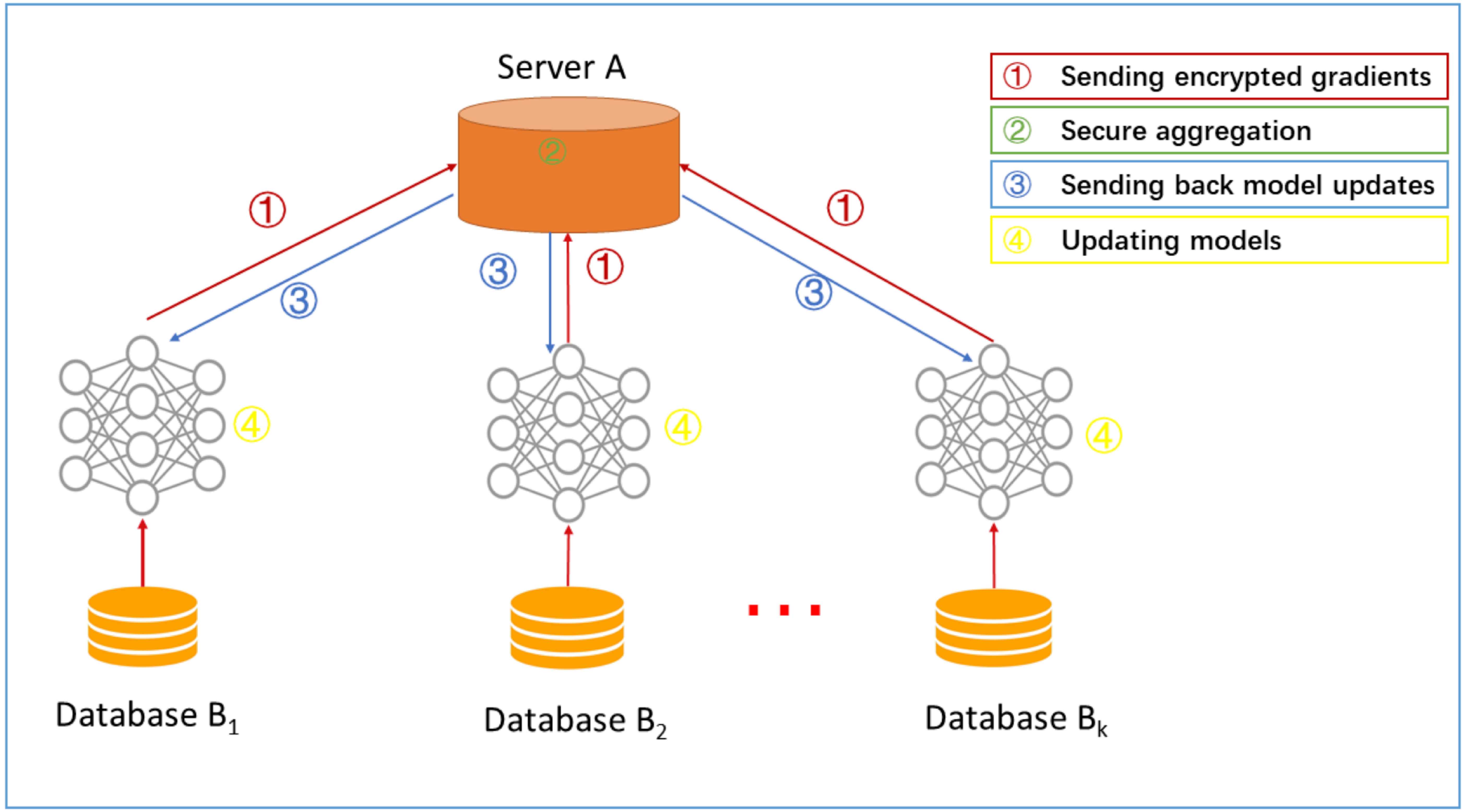
"Lifelong Machine Learning", Chen & Liu, Morgan Claypool, 2017

Much still to be investigated & connected, even beyond the topics we have explored in the course

In retrospect: is data & task heterogeneity at the center?

A slightly different example

Federated learning: different data bases & local "client" models, trained in parallel/with synchronization steps



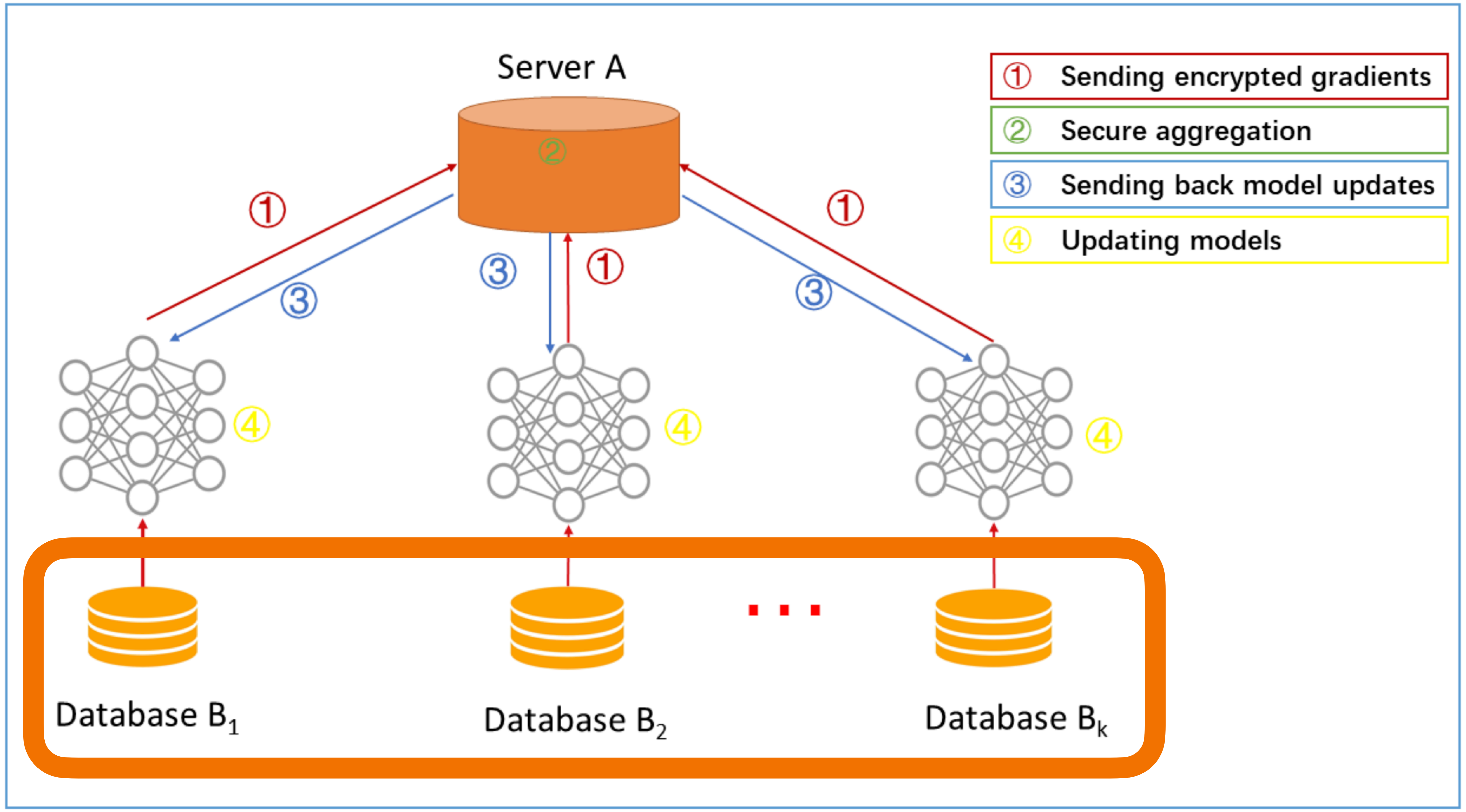
(Some) factors to consider:

- #clients/models
- #updates
- #communication rounds

Figure from "Federated Machine Learning: Concept and Applications", Qiang Yang et al., ACM Journal (TIST), 2019

A slightly different example

We can ask ourselves the same questions again:
what if database distributions/tasks are different?



(Some) factors to consider:

- #clients/models
- #updates
- #communication rounds

Figure from "Federated Machine Learning: Concept and Applications", Qiang Yang et al., ACM Journal (TIST), 2019

Data drift & federated learning

- Horizontally partitioned federated learning (HFL): data distributed in different silos contain the same feature space and different samples
- Vertically partitioned federated learning (VFL): data distributed in different silos contain different feature spaces and the same samples.
- Federated transfer learning (FTL): data distributed in different silos contain different feature spaces and different samples.

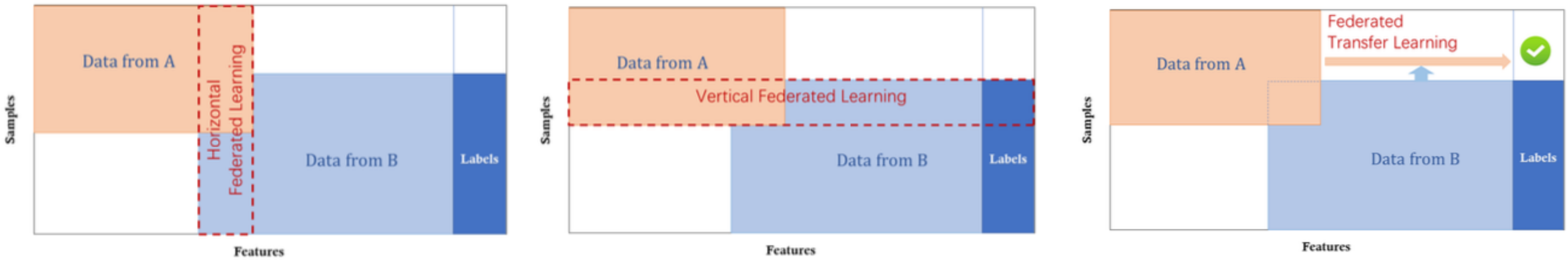
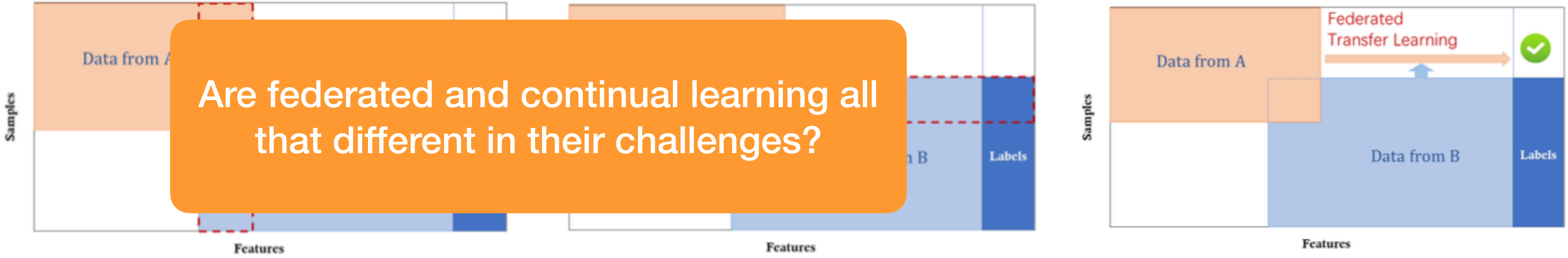


Figure from "Federated Machine Learning: Concept and Applications", Qiang Yang et al., ACM Journal (TIST), 2019

Data drift & federated learning

- Horizontally partitioned federated learning (HFL): data distributed in different silos contain the same feature space and different samples
- Vertically partitioned federated learning (VFL): data distributed in different silos contain different feature spaces and the same samples.
- Federated transfer learning (FTL): data distributed in different silos contain different feature spaces and different samples.



Are federated and continual learning all that different in their challenges?

Figure from "Federated Machine Learning: Concept and Applications", Qiang Yang et al., ACM Journal (TIST), 2019

Continual & federated learning

We can easily think of scenarios where federated + continual go hand in hand

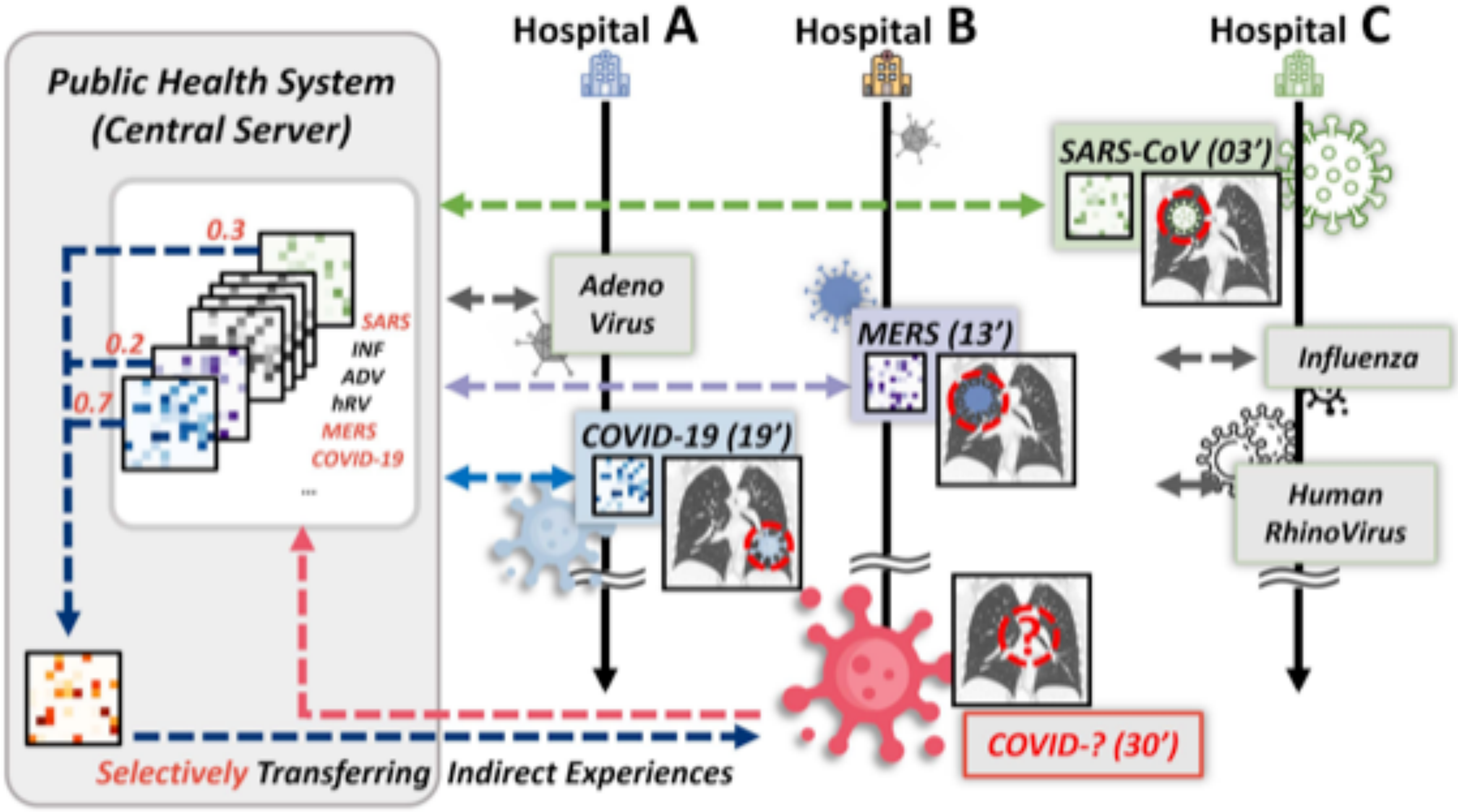


Figure from “Federated Continual Learning with Weighted Inter-client Transfer”, Yoon et al., ICML 2022

Federated continual learning

And then we can start asking ourselves all the same (& more) questions again =)

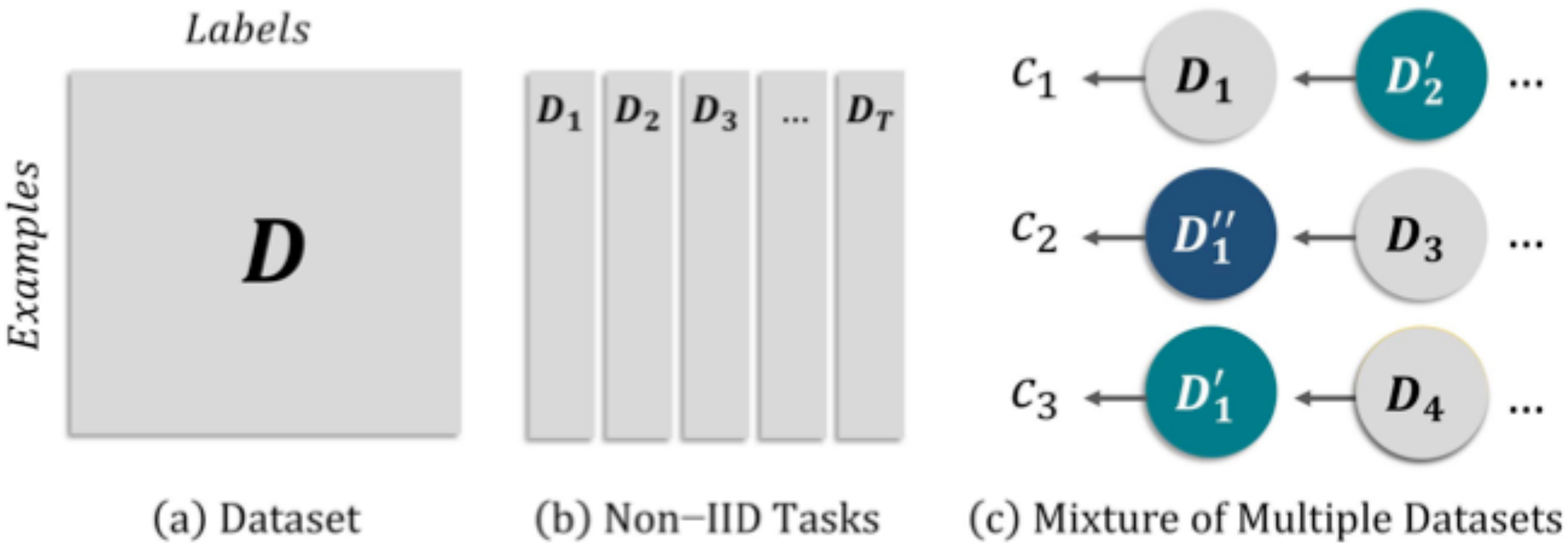
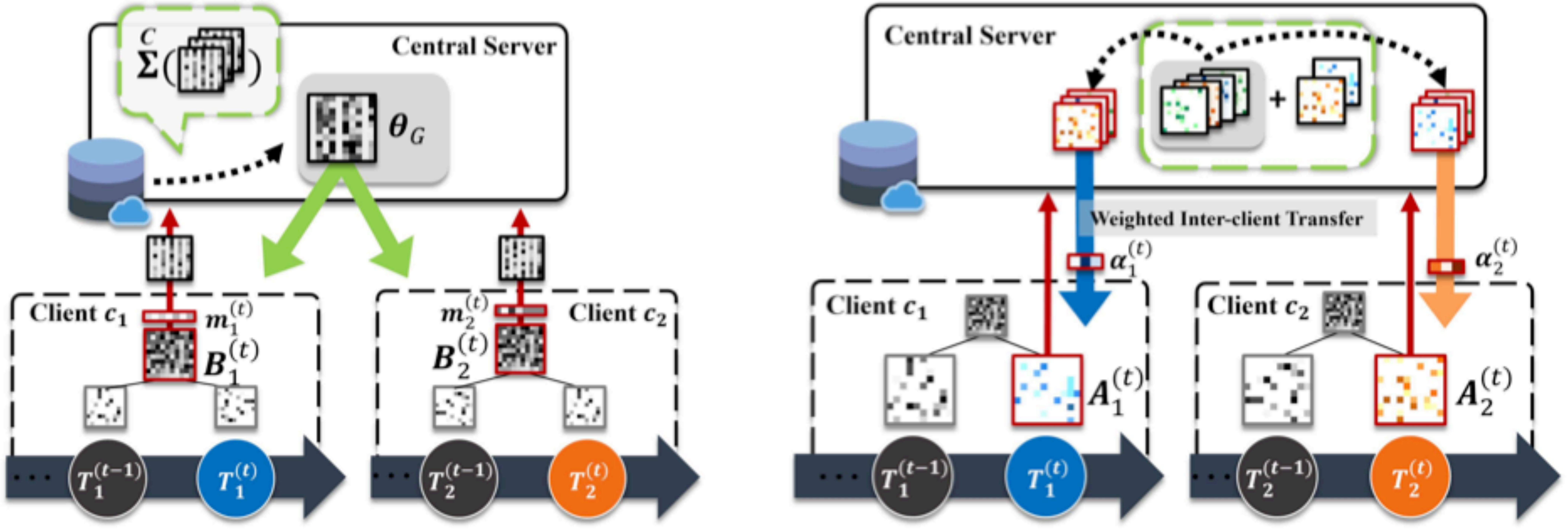


Figure from “Federated Continual Learning with Weighted Inter-client Transfer”, Yoon et al., ICML 2022

Federated continual learning

And we can start applying what we've learned with respect to modular architectures etc.



(a) Communication of General Knowledge (b) Communication of Task-adaptive Knowledge

Figure from "Federated Continual Learning with Weighted Inter-client Transfer", Yoon et al., ICML 2022

Federated continual learning

Perhaps now with other/more trade-offs in mind as well, such as communication costs!

Client communication:

* Communicates a sparsified/masked base parameter B_t * m_t & task-adaptive A_t
* Naive federated learning communicates C (clients) * theta (params) * R (rounds)
* FedWeIT requires C * (R * B + A)

Server communication:

* aggregates/weighted average of masked base parameters
* broadcasts aggregated params theta_t & task adaptive parameters for t-1: A_t-1
* Naive federated learning communicates C * R * theta
* FedWeIT requires C * (R * theta + (C-1)*A) (small overhead of sparse A)

Federated continual learning

Perhaps now with other/more trade-offs in mind as well, such as communication costs!

Client commu

* Communica ... ve A_t
* Naive feder ... nds)
* FedWeIT re

We are back to our question of evaluation & assumptions.

It's perhaps hard to single out a single set of "valid" assumptions & ways to evaluate.

But we do know that it's more than just a single number & a simple train-val-test split!

Server commu

* aggregates/weighted average of masked base parameters
* broadcasts aggregated params theta_t & task adaptive parameters for t-1: A_t-1
* Naive federated learning communicates C * R * theta
* FedWeIT requires C * (R * theta + (C-1)*A) (small overhead of sparse A)

There are so many many more frontiers we don’t have enough time to talk about

Combining even more perspectives, e.g. meta- or online learning, algorithmic/system solutions that are supervision agnostic, important topics such as causality (rather than just correlations)…

The final frontier?

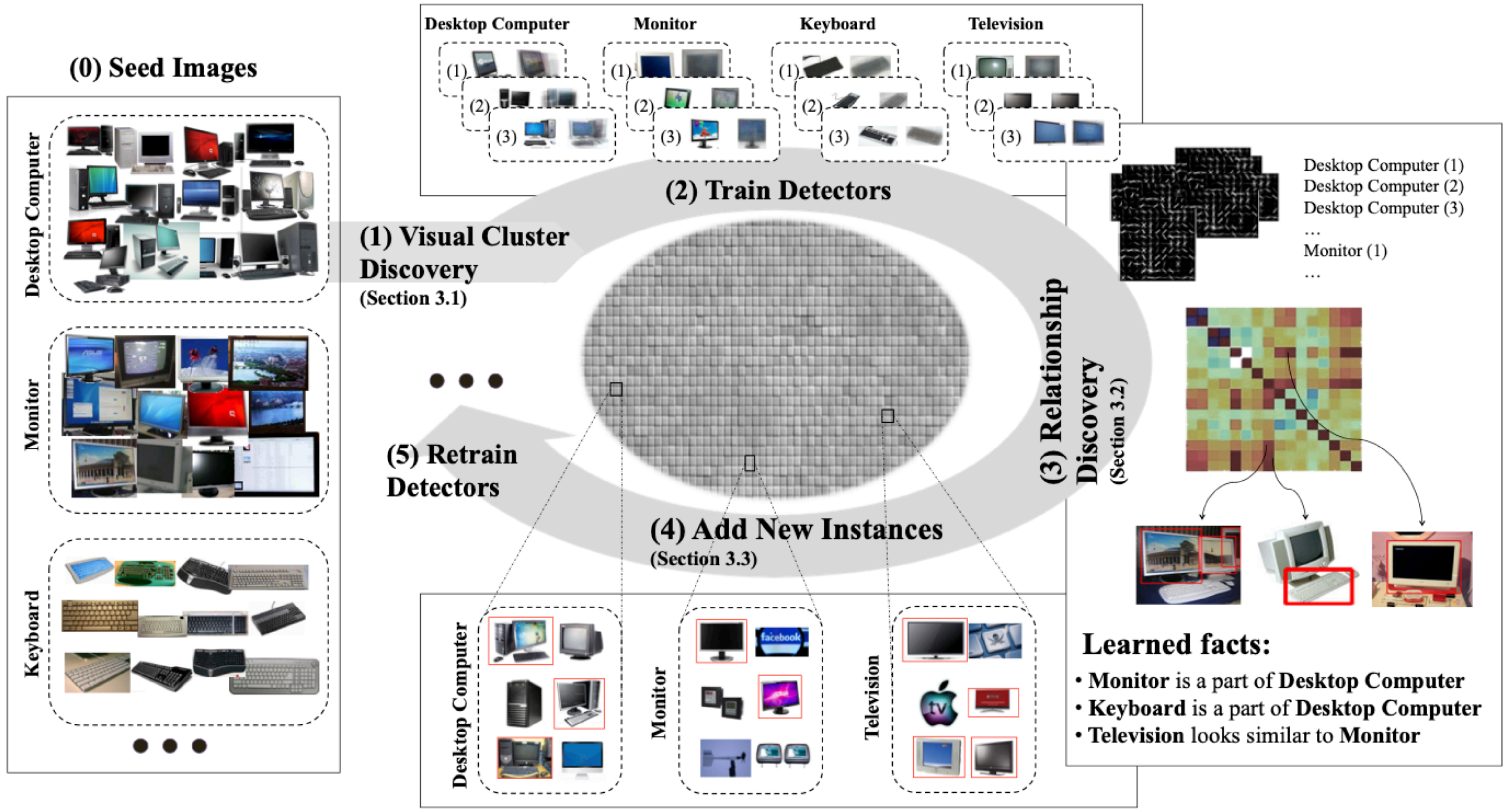
Lifelong open world machine learning?

The final frontier?

Lifelong open world ~~machine learning~~ hybrid AI?

Knowledge, ML & AI

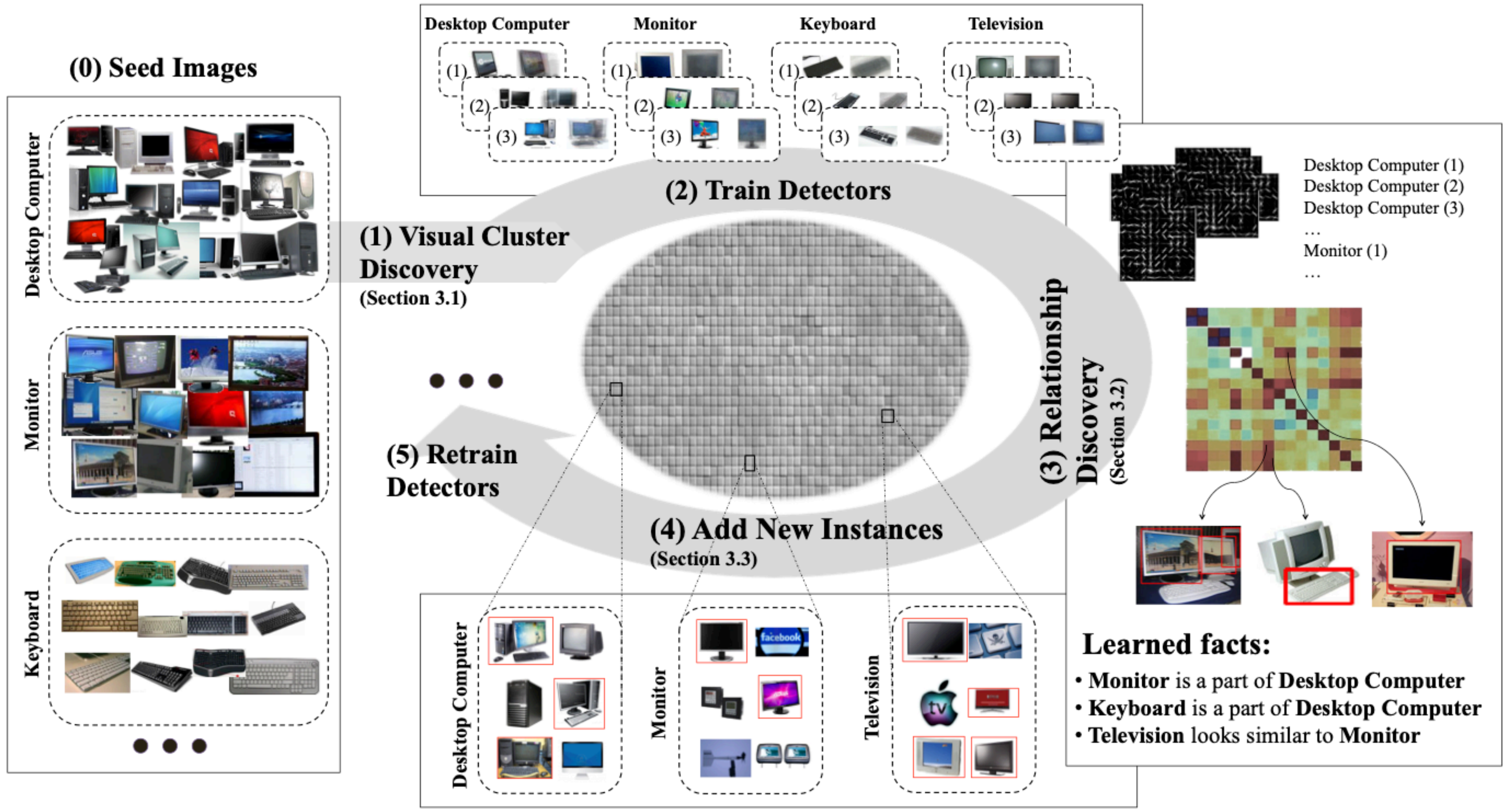
Knowledge is a lot more than just parameters.



"NEIL: Extracting Visual Knowledge form Web Data", X. Chen et al, ICCV 2013

Knowledge, ML & AI

Knowledge is a lot more than just parameters.



NEIL can extract:

- Object categories with bounding boxes
- Labeled examples of scenes
- Examples of attributes
- Visual subclasses of object categories
- Common sense relationships

"NEIL: Extracting Visual Knowledge form Web Data", X. Chen et al, ICCV 2013

Knowledge, ML & AI

Knowledge is a lot more than just parameters.

“We define visual knowledge as any information that can be useful for improving vision tasks such as image understanding and object/scene recognition.

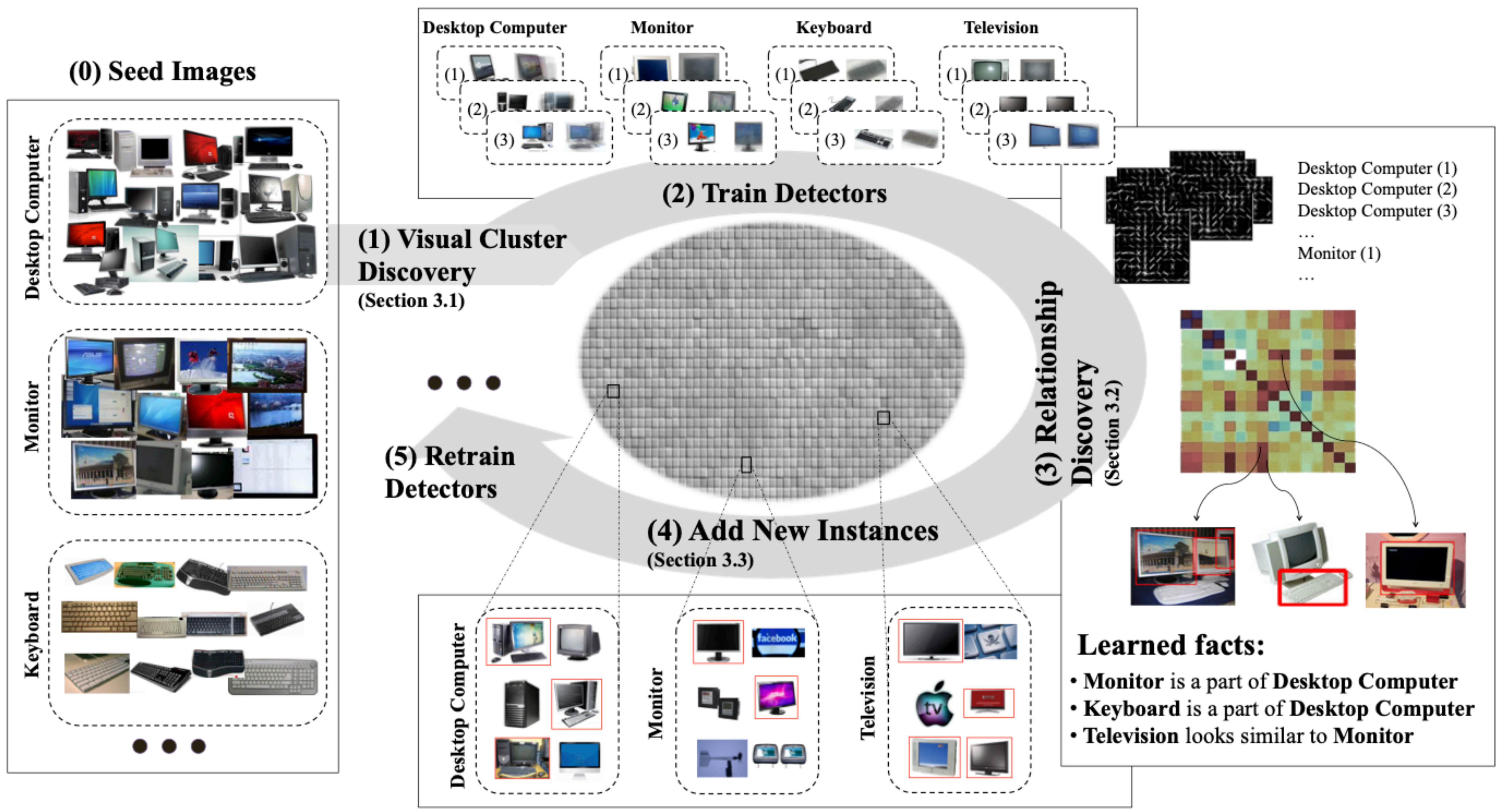
One form of visual knowledge would be labeled examples of different categories or labeled segments/boundaries. Another example would be relationships.

Our knowledge base consists of labeled examples of
(1) Objects; (2) Scenes; (3) Attributes & relationships of 4 types:
(1) Object-Object; (2) Object-Attribute; (3) Scene-Object; (4) Scene-Attribute”

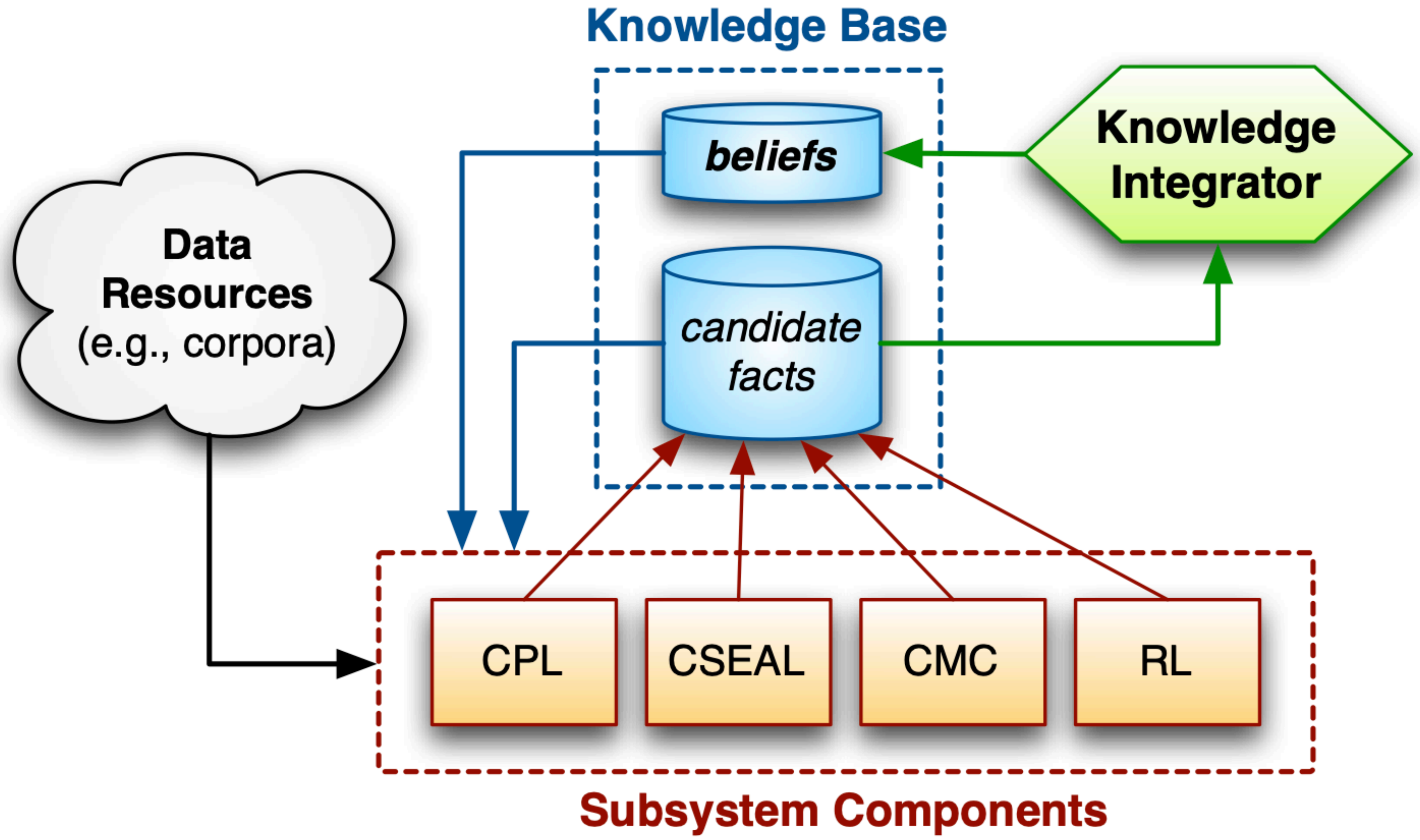
Knowledge, ML & AI

Knowledge is a lot more than just parameters.

AI is more than machine learning!



"NEIL: Extracting Visual Knowledge form Web Data", X. Chen et al, ICCV 2013

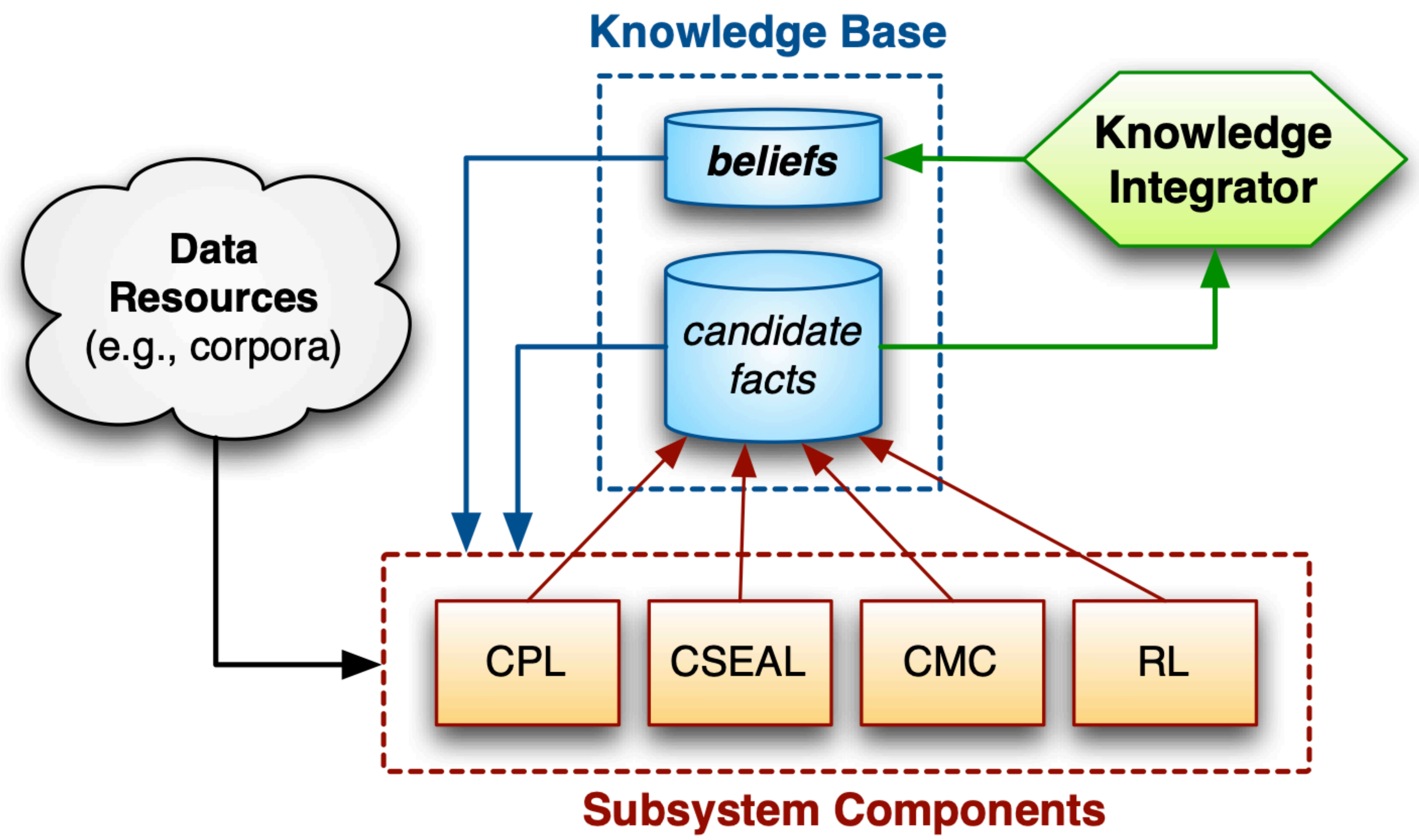


"Towards an Architecture for Never-Ending Language Learning", Carlson et al, AAAI 2010; "Never-Ending Learning", T. Mitchell et al, AAAI 2015

Knowledge, ML & AI

Knowledge is a lot more than just parameters.

AI is more than machine learning!



“Towards an Architecture for Never-Ending Language Learning”, Carlson et al, AAAI 2010; “Never-Ending Learning”, T. Mitchell et al, AAAI 2015

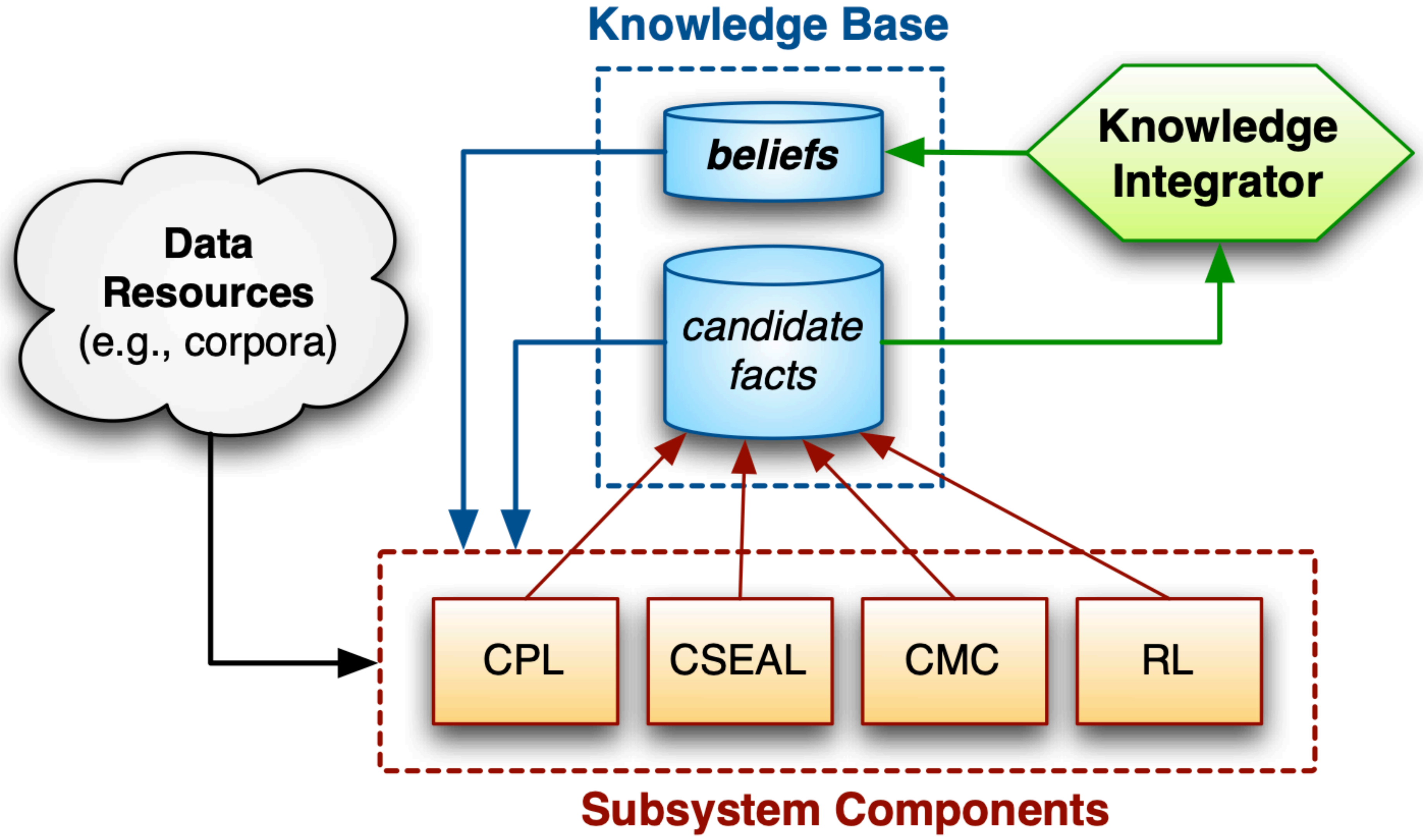
NELL consists of (a really brief overview):

- Coupled Pattern Learner (CPL)
- Coupled Set Expander for Any Language (CSEAL)
- Coupled Morphological Classifier (CMC)
- Rule Learner (RL)
- Knowledge Integrator (KI)
- + NEIL for images (in the second version)

Knowledge, ML & AI

Knowledge is a lot more than just parameters.

AI is more than machine learning!



“Towards an Architecture for Never-Ending Language Learning”, Carlson et al, AAAI 2010; “Never-Ending Learning”, T. Mitchell et al, AAAI 2015

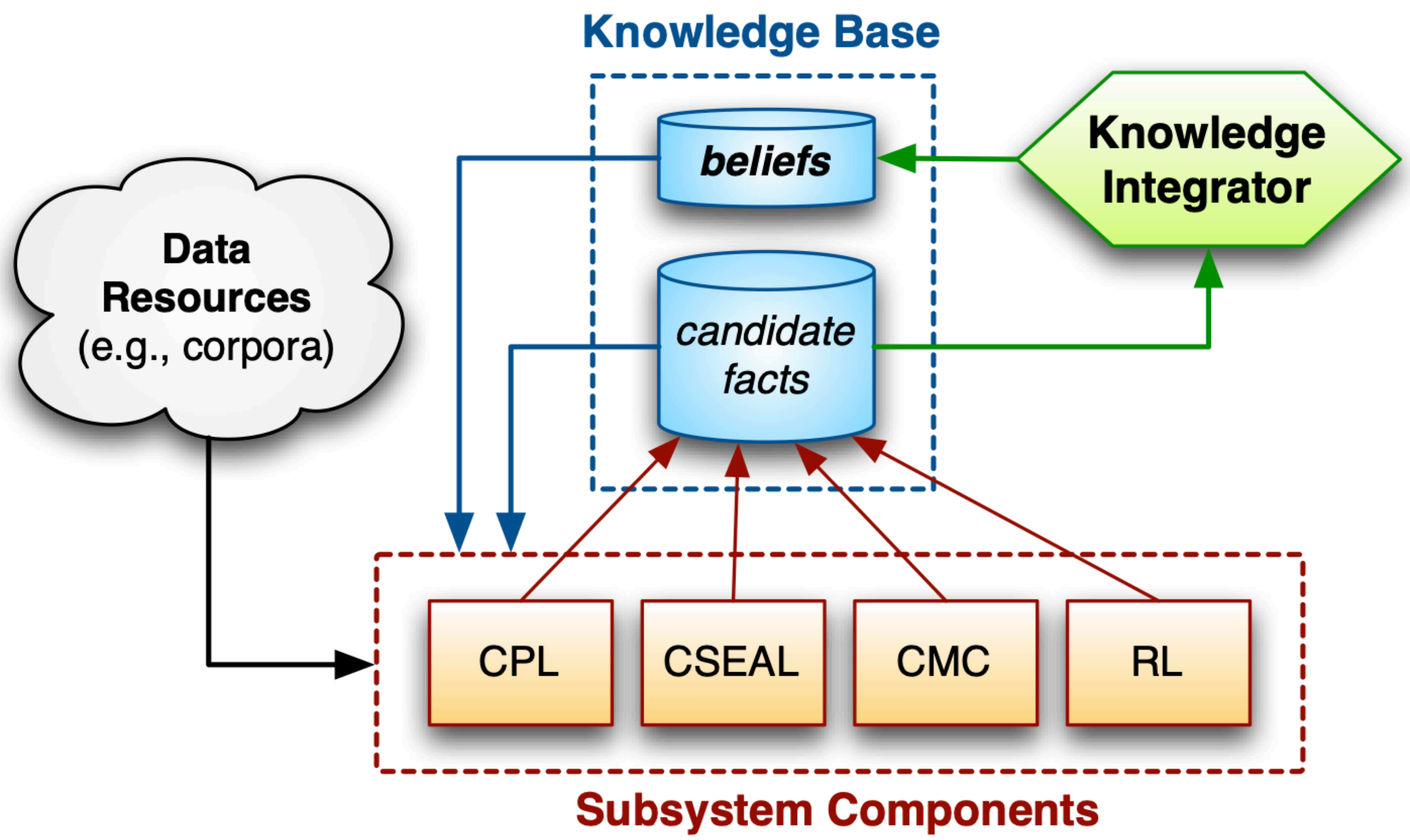
Coupled Pattern Learner (CPL):

- Learns contextual patterns like “mayor of X” and “X plays for Y” to extract categories/relations
- Uses co-occurrence statistics between noun-phrases and contextual patterns
- Relationships are used to filter out patterns that are too general

Knowledge, ML & AI

Knowledge is a lot more than just parameters.

AI is more than machine learning!



"Towards an Architecture for Never-Ending Language Learning", Carlson et al, AAAI 2010; "Never-Ending Learning", T. Mitchell et al, AAAI 2015

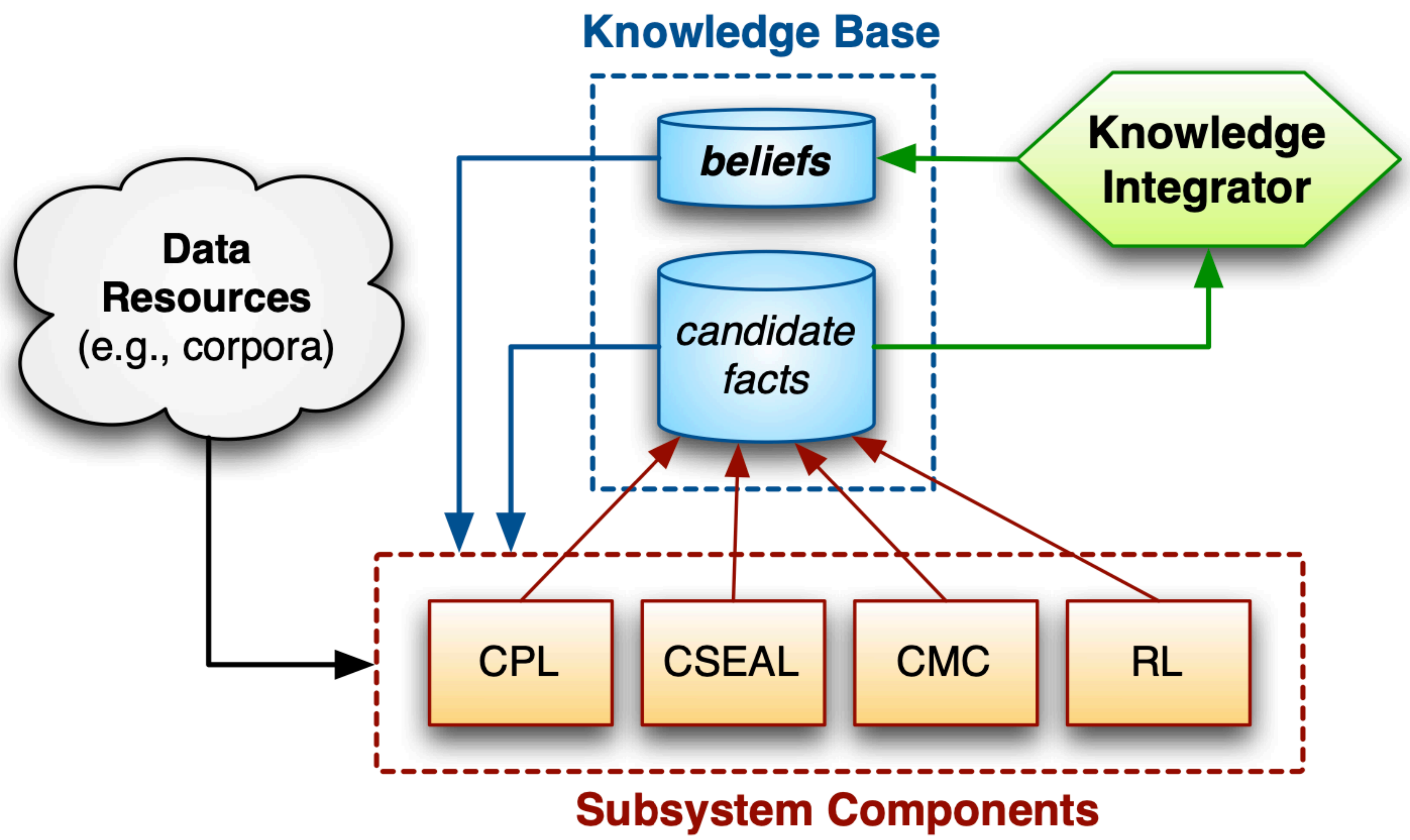
Coupled Set Expander for Any Language (CSEAL):

- Queries internet with sets of beliefs from categories/relations + mines list & tables to extract novel instances
- Uses mutual exclusion relationships to provide negative examples, used to filter out overly general lists and tables

Knowledge, ML & AI

Knowledge is a lot more than just parameters.

AI is more than machine learning!



"Towards an Architecture for Never-Ending Language Learning", Carlson et al, AAAI 2010; "Never-Ending Learning", T. Mitchell et al, AAAI 2015

Coupled Morphological Classifier (CMC):

- Set of binary logistic regression models to classify noun phrases based on morphological features (words, affixes, capitalization, part-of-speech …)

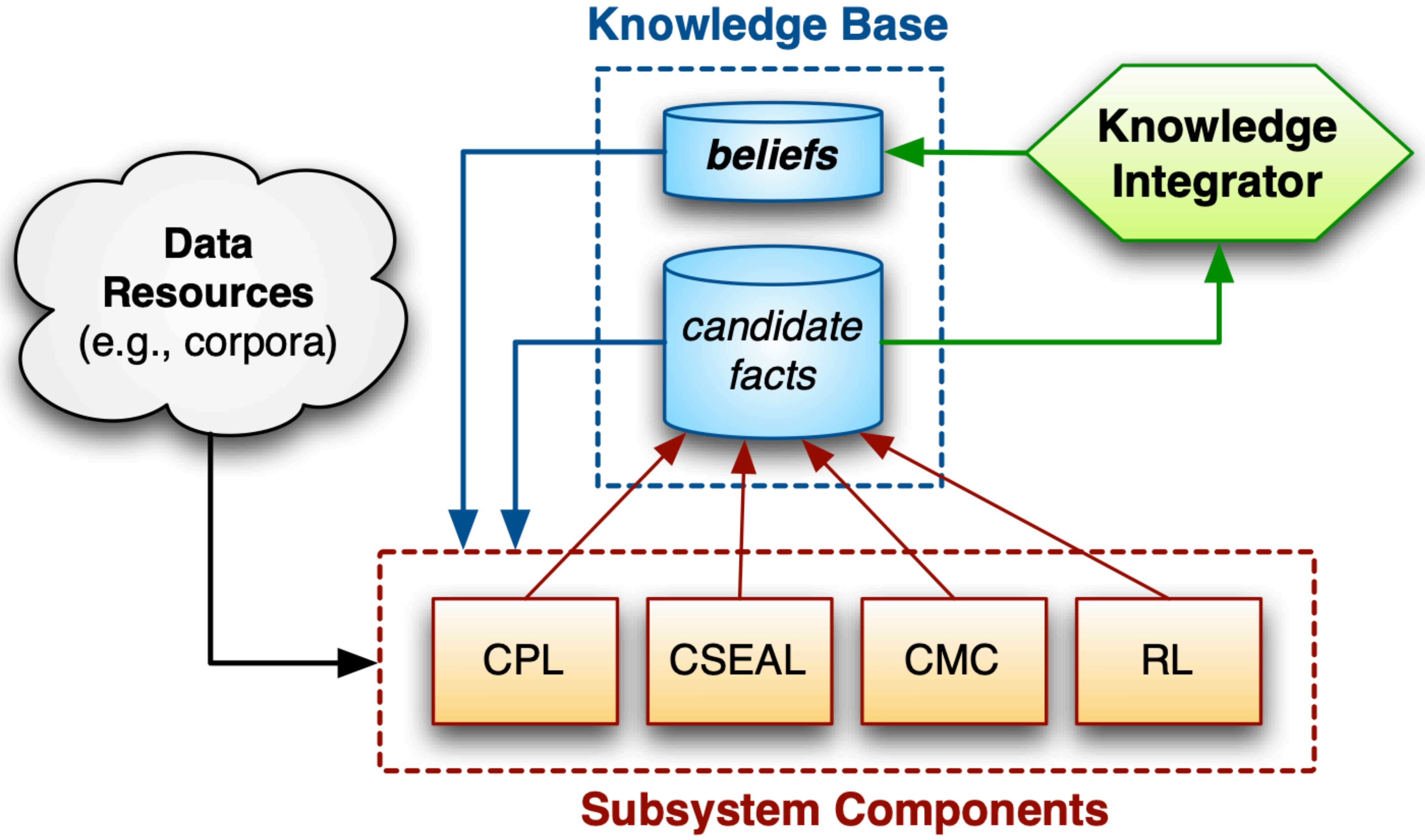
Rule Learner (RL):

- First order relational learning to learn probabilistic Horn clauses. Used to infer new relation instances from other relation instances in the KB

Knowledge, ML & AI

Knowledge is a lot more than just parameters.

AI is more than machine learning!



"Towards an Architecture for Never-Ending Language Learning", Carlson et al, AAAI 2010; "Never-Ending Learning", T. Mitchell et al, AAAI 2015

Knowledge Integrator (KI) + coupling constraints

- Confidence from a single source > 0.9
- Moderate confidence if alternate classifiers agree
- Respects mutual exclusion (disjoint categories)
- Subsets/supersets are coupled & Horn clause coupling (learned mappings are consistent)
- Once promoted/included, never demoted

Keep on learning!

"We will never truly understand machine or human learning until we can build computer programs that, like people,

- *learn many different types of knowledge or functions,*
- *from years of diverse mostly self-supervised experience,*
- *in a staged curricular fashion, where previously learned knowledge enables learning further types of knowledge,*
- *Where self-reflection and the ability to formulate new representations and new learning tasks enable the learner to avoid stagnation and performance plateaus."*

(Quote form the NELL paper, Mitchell et al, AAAI 2015)

Thanks for joining the course!

Q&A session time