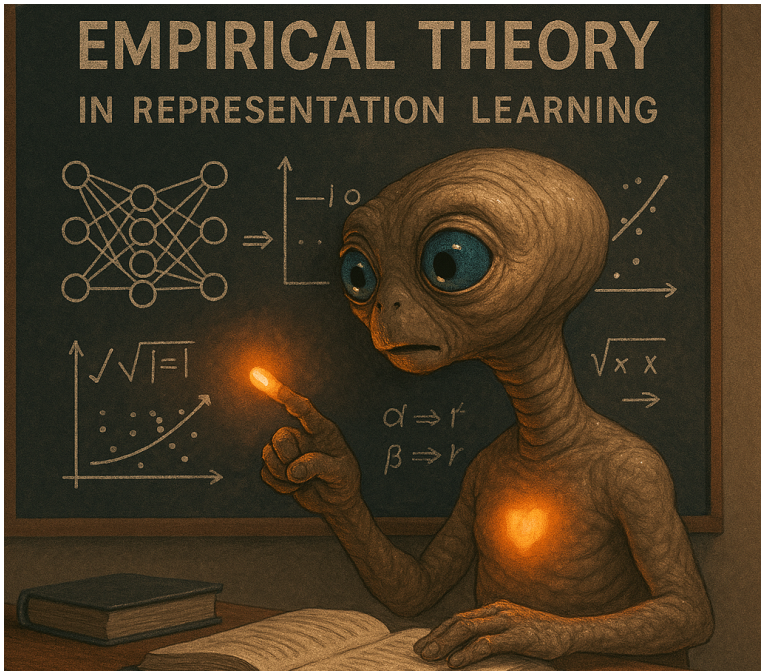
E.T.: Empirical Theory in Representation Learning

A brainstorm about the (under-review) upcoming international workshop at ECCV'26

Jan van Gemert, MS4ML, April 10, 2026



Motivation for an international workshop

- Representation learning: DL methods + large datasets
- Workshop asks: what is "scientific" in representation learning?
- How different from other fields? (evolution, relativity, dual choice theory, ...)

Motivation for an international workshop

- Representation learning: DL methods + large datasets
- Workshop asks: what is "scientific" in representation learning?
- How different from other fields? (evolution, relativity, dual choice theory, ...)

Workshop Goals:

- promote empirical theory research in representation learning
- what is empirical theory?
- What empirical theories do we (implicitly) have?
- How can we encourage rigorous empirical research methods?
- How can we build empirical theories?

What is *Theory*?

- The term *Theory* in ML refers to math-only. (No free lunch, bounds, universal approximation, VC-theory, Rademacher complexity, ...)

What is *Theory*?

- The term *Theory* in ML refers to math-only. (No free lunch, bounds, universal approximation, VC-theory, Rademacher complexity, ...)

Representation learning is practical:

- Exponential Hyper-parameter search important; Non-convex, thus messy.
- Precise, rigorous, theoretical optimality is found as too restrictive and seen as a practically unnecessary artificial hurdle.
- Optimality is replaced with empirical 'good enough'

What is *Theory*?

- The term *Theory* in ML refers to math-only. (No free lunch, bounds, universal approximation, VC-theory, Rademacher complexity, ...)

Representation learning is practical:

- Exponential Hyper-parameter search important; Non-convex, thus messy.
- Precise, rigorous, theoretical optimality is found as too restrictive and seen as a practically unnecessary artificial hurdle.
- Optimality is replaced with empirical 'good enough'

Representation learning is not mathematical; it's empirical

Empirical evidence in representation learning

- Benchmark breaking (bold number) is valuable empirical existential evidence
- Often not clear where the improvements originate: lacks empirical theoretical hypotheses: a clear, causal, link to why it improves
- Ie: If better HP tuning is the reason, this does not increase understanding (we already knew that)
- Instead: *method* for finding HPs generalizes to other works, (and to which works not) and how can this hypothesis be empirically justified?

Empirical evidence in representation learning

- Benchmark breaking (bold number) is valuable empirical existential evidence
- Often not clear where the improvements originate: lacks empirical theoretical hypotheses: a clear, causal, link to why it improves
- Ie: If better HP tuning is the reason, this does not increase understanding (we already knew that)
- Instead: *method* for finding HPs generalizes to other works, (and to which works not) and how can this hypothesis be empirically justified?

Goal of the workshop

- Go beyond individual systems that work well
- Aim for empirical theory: findings that generalize beyond idiosyncratic combinations of datasets, hyper-parameter settings and accidental optimization minima
- Promote hypothesis-driven empirical research that gives insight, and breaking SOTA is neither sufficient nor necessary

Empirical evidence in representation learning

- Benchmark breaking (bold number) is valuable empirical existential evidence
- Often not clear where the improvements originate: lacks empirical theoretical hypotheses: a clear, causal, link to why it improves
- Ie: If better HP tuning is the reason, this does not increase understanding (we already knew that)
- Instead: *method* for finding HPs generalizes to other works, (and to which works not) and how can this hypothesis be empirically justified?

Goal of the workshop

- Go beyond individual systems that work well
- Aim for empirical theory: findings that generalize beyond idiosyncratic combinations of datasets, hyper-parameter settings and accidental optimization minima
- Promote hypothesis-driven empirical research that gives insight, and breaking SOTA is neither sufficient nor necessary

Do research that aims to end up in text books used in teaching.

Objective: Understanding-based representation learning

Aim for a sweet spot between theoretical mathematical models on one side, and purely empirical benchmark-breaking systems on the other side

Objective: Understanding-based representation learning

Aim for a sweet spot between theoretical mathematical models on one side, and purely empirical benchmark-breaking systems on the other side

- Tracing the sources of empirical gain
- Explicitly providing experimental, hypothesis-driven, empirical evidence that separates explanation from speculation
- Understanding the training and the evaluating of deep learning models, their design and their components, optimizers, losses, and how this generalizes over problem and datasets types
- Understanding *when* and *what* makes a method applicable in other work

Objective: Understanding-based representation learning

Aim for a sweet spot between theoretical mathematical models on one side, and purely empirical benchmark-breaking systems on the other side

- Tracing the sources of empirical gain
- Explicitly providing experimental, hypothesis-driven, empirical evidence that separates explanation from speculation
- Understanding the training and the evaluating of deep learning models, their design and their components, optimizers, losses, and how this generalizes over problem and datasets types
- Understanding *when* and *what* makes a method applicable in other work

Examples:

- Is CNN shift invariant? what do residual connections do? Do transformers need registers? What does gating do?
- Neural scaling laws; the lottery ticket hypothesis; the Platonic representation hypothesis

Objective: Understanding-based representation learning

Aim for a sweet spot between theoretical mathematical models on one side, and purely empirical benchmark-breaking systems on the other side

- Tracing the sources of empirical gain
- Explicitly providing experimental, hypothesis-driven, empirical evidence that separates explanation from speculation
- Understanding the training and the evaluating of deep learning models, their design and their components, optimizers, losses, and how this generalizes over problem and datasets types
- Understanding *when* and *what* makes a method applicable in other work

Examples:

- Is CNN shift invariant? what do residual connections do? Do transformers need registers? What does gating do?
- Neural scaling laws; the lottery ticket hypothesis; the Platonic representation hypothesis

Aiming to incentivize empirical insights, and empirical theory building.

Conference-style pre-registration initiative

Pre-registration is successful in other fields; with two goal:

1. Avoid spurious correlations; 'P-hacking'
2. Accept RQ + method independent of pos/neg findings.

Conference-style pre-registration initiative

Pre-registration is successful in other fields; with two goal:

1. Avoid spurious correlations; 'P-hacking'
2. Accept RQ + method independent of pos/neg findings.

Difficult to do in ML: we have conferences, with fixed, short, deadlines.

Conference-style pre-registration initiative

Pre-registration is successful in other fields; with two goal:

1. Avoid spurious correlations; 'P-hacking'
2. Accept RQ + method independent of pos/neg findings.

Difficult to do in ML: we have conferences, with fixed, short, deadlines.

Trial: A submission needs both a (possibly pre-registered) storyline and a paper.

Conference-style pre-registration initiative

Pre-registration is successful in other fields; with two goal:

1. Avoid spurious correlations; 'P-hacking'
2. Accept RQ + method independent of pos/neg findings.

Difficult to do in ML: we have conferences, with fixed, short, deadlines.

Trial: A submission needs both a (possibly pre-registered) storyline and a paper.

Storyline (possibly pre-registered):

- Storyline of 1 page, templated, research proposal
- Reviewed, given feedback, and scored on acceptance likelihood if deemed interesting, independent of findings.

Paper

- Review as usual; clarity, related work, methodology; experimental setup
- Check if paper aligns with storyline. Cannot reject based on findings.

Conference-style pre-registration initiative

Pre-registration is successful in other fields; with two goal:

1. Avoid spurious correlations; 'P-hacking'
2. Accept RQ + method independent of pos/neg findings.

Difficult to do in ML: we have conferences, with fixed, short, deadlines.

Trial: A submission needs both a (possibly pre-registered) storyline and a paper.

Storyline (possibly pre-registered):

- Storyline of 1 page, templated, research proposal
- Reviewed, given feedback, and scored on acceptance likelihood if deemed interesting, independent of findings.

Paper

- Review as usual; clarity, related work, methodology; experimental setup
- Check if paper aligns with storyline. Cannot reject based on findings.

The storyline can be submitted any time; and is given feedback, time permitting. The paper has a final deadline. If the storyline was submitted too late, it is judged at the same time as the paper.

Example review questions

- What is interesting about the paper?
- Explanation vs Speculation: are all claims supported with empirical evidence?
- Sources of empirical gains: what empirical evidence is there to support that the improved accuracy comes from what is claimed, and not due to some other confounding effect? (eg: hyper-parameters?)
- Methodological generalizability: how are the findings relevant to other methods/papers?

Relation to other workshops and initiatives

Great examples to learn from:

- Metascience for Machine Learning :)
- Pre-registration workshop: An alternative publication model for machine learning research; NeurIPS 2011
- The ML-Retrospectives, Surveys & Meta-Analyses workshop, NeurIPS 2019, ICML 2020, NeurIPS 2020
- Mechanistic Interpretability Workshop, ICML 2024 and NeurIPS 2025
- Workshop on Scientific Methods for Understanding Deep Learning, ICLR 2025/2026

We here aim to foster, and build a community for understanding-based research that is currently scattered over multiple venues, and mixed in with improvement-based research.

Organizers

Jan van Gemert

Hayley Hung

Marco Loog

Ana Lucic

David Picard

Hannah Pinson

Questions; feedback; criticism?

- What is 'theory' anyway?
- Pre-registration goals?
- Is a new venue needed at all?