

Marek Rosa - A Framework for Searching for General Artificial Intelligence (2016)

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — [861fb9d0](#)

0.1 Context

0.2 Learned in this study

- Problems we'd like to solve
 - Gradual learning
 - Incremental learning
 - Guided learning

0.3 Things to explore

- How do they plan to solve the “teaching” problem (communication between teacher and agent)?
- Should curricula “learning” be developed in a genetic algorithm fashion (lots of tested curricula, the best one survive and go on to be adopted by the next batch of agents)

1 Overview

- Gradual learning is the idea that you slowly increase the number of parameters that a model is allowed to use to describe a task

2 Notes

2.1 Abstract

- Intelligence is a tool for searching for general solutions to problems
- Intelligence is the ability to acquire skills that narrow this search, diversify it and help steer it to more promising areas (heuristics?)

2.2 Introduction

- We view artificial intelligence (AI) systems as programs that are able to learn, adapt, be creative and solve problems

2.3 Foundations of general AI

2.3.1 What is intelligence?

- We view intelligence as a problem solving tool that searches for solutions to problems in dynamic, complex and uncertain environments

- From a computational point of view, most solvable problems can be viewed as search and optimization problems, and the goal of intelligence (or an intelligent agent) is to always find the best available solutions with as few resources and as quickly as possible
- Intelligence can achieve this by discovering skills (abilities, heuristics, shortcuts, tricks) that narrow the search space, diversify it, and efficiently help steer it towards areas that are potentially more promising
- Some of the most useful skills are the capacity to gradually acquire new skills - which helps in exploiting accumulated knowledge in order to speed up the acquisition of additional skills, the reuse of existing skills, and recursive self-improvement
- The possibly limitless accumulation of skills in an intelligent machine is naturally bounded by limited resources (time, memory, atoms, energy, etc.). This additional constraint on intelligence results in favoring “efficient” skills and operation under bounded rationality

2.3.2 What is a skill?

- A skill is any assumption about a problem that narrows and diversifies the search for a solution and points the search towards more promising areas
- Synonyms: ability, heuristic, behavior, strategy, solution, algorithm, shortcut, trick, approximation, exploiting structure in data, and more
- A skill can also be considered a bias. It constrains the search space or restricts behavior
- One possible way to compare the intelligence of various agents is to measure and compare the complexity and generality of problems they can solve
- Another way can be in terms of measuring the effectiveness as well as efficiency of their skills in terms of their sample complexity, the required computational cycles or their reuse of previously learned skills

2.4 How to build and educate general AI?

2.4.1 How to optimize the process?

- We can start asking questions like, “What is the minimal skill set that is sufficient for human-level general AI?”
- How is our approach better compared to others?
- Is the notion of skill sufficiently well defined?
- Is the ordering structure of our curricula sufficiently rich?

2.5 The importance of a roadmap to general artificial intelligence

- The roadmap is partitioned into two primary areas, containing three complementary parts
 - Research and Development - how to get to general AI?
 - * Architecture Roadmap - intrinsic skills and architecture design
 - * Curriculum Roadmap - learned skills and gradual knowledge acquisition
 - Future and Safety - what to expect next?
 - * Safety/Futuristic Roadmap - how to keep humanity safe and the years leading up to and after general AI is reached
- To learn general skills, we can define what tasks the agent should be able to solve. Based on tasks that have not been solved yet, we can derive a research problem (what researchers should figure out). Solving this research problem results in at least one of the following:
 - A modified architecture which can solve the task (exhibits new intrinsic skill(s))
 - A modified architecture which can learn how to solve the task (exhibits new intrinsic and/or learned skill(s))
 - A modified curriculum, in which the system can learn to solve tasks (acquired new learned skill(s))

2.5.1 Guidelines for working with the roadmap

- The following key questions are essential for building successful curricula:
 - How to come up with problems that will appear in the curriculum?

- How to determine the right time to present a problem in the curriculum?
- How to decide that a problem is useful / not useful or is missing?
- How to decide that two problems are similar / different?
- How to correctly evaluate an agent's solution to a problem?

2.6 AI Roadmap Institute

~

2.7 School for AI/Curriculum learning

- Beside having hard-coded general intrinsic skills, the system is expected to be able to learn
- First we design an optimized set of learning tasks, or a “curriculum”
- The aim of the curriculum is to teach the AI system useful skills and abilities, so it does not have to discover them on its own

2.7.1 Curriculum requirements

- A good curriculum:
 - Minimizes the time needed for getting the general AI into a desired target state. When the AI is in a target state, it can learn and evolve on its own
 - Is efficient - i.e. not more complex than necessary
 - Minimizes the number of skills that need to be hard-coded into the general AI

2.7.2 Artificial learning environment

- During development of the School for AI, an interesting problem was encountered - how should the tasks be specified for the AI?
- When there is little or no common language, it is very challenging and time-consuming to explain what the AI needs to do
- For this reason, focus is on early language acquisition
- To cut down on AI development time, we want to be able to efficiently communicate with it as soon as possible

2.8 Gradual learning competition

~

2.9 Development of theoretical foundations

- Information Theory
 - Algorithmic Information Theory
 - Kolmogorov Complexity
 - Minimum Description/Message Length
- Learning Theory
 - Vapnik-Chervonenkis Theory
 - Rademacher Complexity
 - Robust Generalization
- Computational Mechanics and Statistical Physics
 - Structural complexity
 - ϵ -machines and transducers
 - Integrated Information

2.10 Related work

- Mikolov et al., 2015: Focuses on communication and learning in a simplified “toy” environment
- Lake et al., 2016: A more philosophical discussion of the limitations of current approaches. They argue that solving pattern recognition is not sufficient and that causal discovery is essential
- Bengio et al., 2009: BabyAI project and introduction of curriculum learning: learning accelerated by presenting easy examples first and progressively increasing the difficulty
- Thórisson et al., 2016: A theory of AI tasks can give us more rigorous ways of comparing and evaluating intelligent behavior
- Flavell et al., 2002: Description of stages of human cognitive development
- Hernández-Orallo et al., 2016: An overview of evaluation environments, measures and challenges in AI
- Of particular interest are growing architectures that learn new things while retaining existing knowledge
- Pan and Yang, 2010: A systematic overview of a related field of “transfer learning”
- Schmidhuber, 2013: A framework for automatically discovering problems inspired by playful behavior in animals and humans
- Rusu and al., 2016: Progressive neural networks that adapt to new tasks by growing new columns
- Fahlman and Lebiere, 1990: Accelerated learning by adding one hidden neuron at the time, keeping the preceding hidden weights frozen
- Compositional learning, i.e. the ability to form knowledge about a particular subject by unifying knowledge about multiple other subjects that are already understood
- Learning programs or algorithms are another form of compositional learning

3 See also

4 References

- Rosa, Marek, Jan Feyereisl, and The GoodAI Collective. “A Framework for Searching for General Artificial Intelligence.” arXiv preprint arXiv:1611.00685 (2016).
- <https://arxiv.org/abs/1611.00685>