
Consolidating the Search for General AI

Marek Rosa* and Jan Feyereisl
GoodAI
Prague, Czech Republic
{marek.rosa, jan.feyereisl}@goodai.com

Abstract

There is a significant lack of unified approaches to building generally intelligent machines. Artificial intelligence research frequently operates within a very narrow field of focus, oftentimes without considering the importance of the “big picture”. In this position paper, we outline our steps towards rectifying this situation. We propose a number of processes that we believe encourage a more comprehensive approach to searching for generally intelligent machines. These include our framework, agent development roadmap, AI Roadmap Institute and a research competition, each playing a different role in consolidating the necessary steps in order to progress in arguably the most significant search of this century.

1 Introduction

In artificial intelligence (AI) research, akin to the natural sciences [11], many researchers, universities and institutes frequently operate within limited boundaries, clearly defined by their narrow focus of interest. Their specialization inherently limits their consideration of the field as a whole and hence hinders progress. Here, we suggest a number of ways to step out of this cycle and provide a unified perspective on building machines that learn to think.

We start with our *framework* [16], a “big picture” view and description of processes, definitions and ideas that we deem essential in our search for general AI. This is followed by the introduction of our *agent development roadmap* [17], a step-by-step guide for the gradual and guided accumulation of skills by an intelligent agent—a vital, yet commonly neglected aspect of agent learning. Then, we introduce the idea of an *AI Roadmap Institute* [19] that keeps track of the progress of the entire field and maps disparate research into a common and easily interpretable representation, encouraging more unified approaches to general AI research. Finally, we propose the creation of a *competition*, seeking the solution to the problem of gradual learning under computational rationality [5]. Additionally, in the Appendix we provide a very brief comparison of our realization of the agent development roadmap, our “School for AI”, and the recently introduced CommAI-env environment [13]. This should serve as a brief sample of the type of analyses at the core of the proposed AI Roadmap Institute.

2 A framework for searching for general AI

In our framework document [16], we seek to describe and unify principles that guide our development of general AI. These principles revolve around the idea that intelligence is a tool for searching for solutions to problems. We define intelligence as the ability to acquire skills that narrow this search, diversify it and help steer it to more promising areas. We provide suggestions for studying, measuring, and testing the various skills and abilities that a machine of human-level intelligence [14] needs to

*The authors would like to thank the entire GoodAI team for their help and support in writing this manuscript and for all the work without which none of this would be possible.

acquire. The framework aims to be implementation agnostic and to provide an analytic, systematic, and scalable way to generate hypotheses that we believe are needed to meet the minimal and necessary conditions in the search for general AI.

The framework also provides a list of “*next steps*”: important research topics that we believe the community as a whole needs to focus on next in order to allow for significant progress in the field. These include a) publishing of “big picture” overviews by other researchers, akin to our framework [16], roadmap [17] and other sporadic related works [13, 22, 10, 15], b) provision of unified theoretical foundations applicable across frameworks, c) development of a task theory [21] measuring complexity of learning tasks, d) continual development of more learning tasks, and e) encouragement of collaboration within the AI Roadmap Institute.

With a modular definition of the problems in our agent development roadmap, the work can be split among various research groups (both internally as well as among external collaborators, academia, students, other research centers). The framework does not focus on narrow artificial intelligence (which solves very specific problems well), or on short-term commercialization. The aim is truly long-term, with possible exploitation of useful applications along the way. The first version of the framework is for both a general and a technical audience. The aim is to make it first accessible to everyone, yet over time with enough detail that advanced readers will also benefit from it. We believe that such a framework is the first stepping stone towards bringing together definitions, highlighting open problems and connecting researchers willing to collaborate.

3 The importance of a roadmap to general AI

Our agent development roadmap [17] is a principled approach to clearly outlining and defining a step-by-step guide for obtaining all skills that a human-level intelligent machine needs to possess. This includes their definitions, as well as the gradual order and way in which to achieve them through curricula of our “School for AI”.

It is a collection of research milestones that we deem essential for progress towards general AI. Currently, we partition the agent development roadmap into the following areas: a) *Architecture Roadmap* - what are the necessary intrinsic (i.e. hard-coded) skills and the necessary architectural design, and b) *Curriculum Roadmap* - what learned skills are required and how to gradually acquire knowledge. Both roadmaps contain partially ordered lists of skills which our AI will need to exhibit in order to achieve human-level intelligence.

A *skill* is an ability of an agent to complete a task (solve a problem in an environment). We can define what tasks the agent should be able to complete. Based on tasks that are not completed, we can derive a research problem (or a milestone). Solving this research problem results in one of the following: a) a modified architecture which can complete the tasks—it exhibits new intrinsic skill(s), b) a modified architecture which can learn how to complete the tasks—it exhibits new intrinsic and/or learned skill(s), or c) a modified curriculum, in which the system can learn to complete the tasks—it has acquired new learned skill(s).

New skills very often depend and build upon previously acquired skills, so the research milestones exhibit some inherent dependencies. We shouldn’t simply skip to a skill in the middle of the roadmap and start acquiring it. This could inhibit the potential efficiency benefits that the reuse of previously learned skills might offer. Instead, each skill should also be a stepping stone to subsequent skills—a fundamental property of learning we call *gradual learning*. It is very important that an architecture that solves problems (tasks) in the roadmap does not approach each problem in isolation. On the contrary, the solution of a problem could ideally be based on the solutions of previous, simpler problems [3, 6]. Under such a graduality requirement, some problems that are “solved” in the traditional sense of the word, like chess or checkers, still remain open. To encourage the community to also engage in roadmapping, we will publish a guideline for working with the agent development roadmap and for creating curricula [1].

4 The AI Roadmap Institute

In an attempt to provide a platform for better collaboration and understanding between researchers and to measure genuine progress in the search for general AI, we propose the creation of an AI

Roadmap Institute. We are founding and starting this new initiative [19], to collate and study various AI roadmaps proposed by those working in the field, map them into a common representation and therefore enable their comparison. The institute will use architecture-agnostic common terminology to compare roadmaps, allowing research groups with different internal terminologies to communicate effectively. The institute is concerned with “big picture” thinking, without unnecessarily focusing on local problems in the search for general AI.

The amount of research into AI has exploded over the last few years, with many new papers appearing daily. The institute’s major output will be consolidating this research into a comprehensible summary which outlines the similarities and differences among roadmaps and which maps progress in the field in general. This summary will identify where roadmaps branch and converge, show stages of roadmaps which need to be addressed by new research, and highlight examples of skills and testable milestones. The roadmaps will show problems and any proposed solutions, and the implementations of others will be mapped out in a similar manner. The summary will be presented in a clear and comprehensible way to maximize its impact on as wide an audience as possible, minimizing the need for significant technical expertise, at least at its “big picture” level. With a point of comparison among different roadmaps and with links to relevant research, the institute can highlight aspects of AI development where solutions exist or are needed. This means that other research groups can take inspiration from or suggest new milestones for the roadmaps. Finally, the institute is for the scientific community and everyone will be invited to contribute. It will be constantly updated and available for all who are interested. In the Appendix, we provide a very brief example of one type of analysis the institute will undertake on a daily basis.

5 Competition: Gradual learning under computational rationality

We believe that one of the most fundamental challenges in developing human-level intelligent machines is the creation of agents that have the ability to acquire and reuse skills and knowledge in a gradual manner. Unlike in an unconstrained setting, this problem continues to pose serious challenges under bounded resources[5]. To truly and quickly progress in this area, we propose the injection of a monetary stimulus to the AI community in the form of a competition. We suggest launching the competition in two stages: 1) *Stage 1* - Identification of requirements, specifications and a set of evaluation tasks for gradual learning, 2) *Stage 2* - Development and implementation of an agent that gradually learns and passes requirements defined in stage 1. Upon completion, the winner will receive a significant monetary prize (millions of \$), provided by us and potentially by other investors.

6 Conclusion

In this position paper, we have proposed a number of ways that we believe could speed up the search for general AI. First, we encourage more researchers to step back and away from their research local optima and consider a broader view, in a similar manner to our framework. Then, we propose an often underestimated holistic view of all stages of an agent’s learning experience, our agent development roadmap, that encourages learning in a gradual and guided fashion. To foster collaboration, encourage further efforts akin to the ones proposed here and to map and measure genuine progress in the field, we describe the idea behind the AI Roadmap Institute that we are founding. Finally, we propose the creation of a prize-driven competition for solving one of the most important challenges in developing general AI—gradual learning under computational rationality. We invite and encourage everyone to take part in both the AI Roadmap Institute and the competition, whether as active members, users, competitors or in any other form possible.

References

- [1] S. Andersson, M. Poliak, M. Stransky, and The GoodAI Collective. Building Curriculum Roadmaps for Artificial Agents - version 1.0. *to be published*, 2016.
- [2] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling. The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research*, 47:253–279, jul 2012.
- [3] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

- [4] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. OpenAI Gym. *arXiv:1606.01540*, pages 1–4, jun 2016.
- [5] S. J. Gershman, E. J. Horvitz, and J. B. Tenenbaum. Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, 349(6245):273–278, 2015.
- [6] Ç. Gülçehre and Y. Bengio. Knowledge Matters: Importance of Prior Information for Optimization. *Journal of Machine Learning Research*, 17(8):1–32, 2016.
- [7] J. Hernández-Orallo. Evaluation in artificial intelligence: from task-oriented to ability-oriented measurement. *Artificial Intelligence Review*, pages 1–51, 2016.
- [8] M. Johnson, K. Hofmann, T. Hutton, and D. Bignell. The Malmo Platform for Artificial Intelligence Experimentation. *International joint conference on artificial intelligence (IJCAI)*, page 4246, 2016.
- [9] M. Kempka, M. Wydmuch, G. Runc, J. Toczek, and W. Jaśkowski. ViZDoom: A Doom-based AI Research Platform for Visual Reinforcement Learning. *arXiv:1605.02097v1*, may 2016.
- [10] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. Building Machines that learn and think like people. *arXiv:1604.00289v1*, pages 1–55, apr 2016.
- [11] H. Ledford. How to solve the world’s biggest problems. *Nature*, pages 1–12, 2015.
- [12] A. Lerer, S. Gross, and R. Fergus. Learning Physical Intuition of Block Towers by Example. *arXiv:1603.01312*, mar 2016.
- [13] T. Mikolov, A. Joulin, and M. Baroni. A Roadmap towards Machine Intelligence. *ArXiv:1511.08130*, pages 1–39, 2015.
- [14] L. Muehlhauser and A. Salamon. Intelligence Explosion : Evidence and Import. *Singularity Hypotheses: A Scientific and Philosophical Assessment*, page 30, 2012.
- [15] E. Nivel, K. R. Thórisson, B. R. Steunebrink, H. Dindo, G. Pezzulo, M. Rodriguez, C. Hernandez, D. Ognibene, J. Schmidhuber, R. Sanz, H. P. Helgason, A. Chella, and G. K. Jonsson. Bounded Recursive Self-Improvement. *arXiv:1312.6764*, dec 2013.
- [16] M. Rosa, J. Feyereisl, and The GoodAI Collective. A Framework for Searching for General Artificial Intelligence - version 1.0. *to be published Nov. 2016*, 2016.
- [17] M. Rosa, M. Poliak, J. Feyereisl, S. Andersson, M. Vlasak, M. Stransky, O. Sota, and The GoodAI Collective. GoodAI Agent Development Roadmap - version 1.0. *to be published Nov. 2016*, 2016.
- [18] S. Sukhbaatar, A. Szlam, G. Synnaeve, S. Chintala, and R. Fergus. MazeBase: A Sandbox for Learning from Games. *arXiv:1511.07401*, pages 1–11, nov 2015.
- [19] The GoodAI Collective. AI Roadmap Institute. *to be launched*, url: <http://www.goodai.com/ai-roadmap-institute>, [Accessed: 08- Oct- 2016], 2016.
- [20] The GoodAI Collective. GoodAI - School for AI. *to be released - open-source*, url: <http://www.goodai.com/school-for-ai>, [Accessed: 08- Oct- 2016], 2016.
- [21] K. R. Thórisson, J. Bieger, T. Thorarensen, J. S. Sigurðardóttir, and B. R. Steunebrink. Why artificial intelligence needs a task theory: And what it might look like. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9782, pages 118–128, apr 2016.
- [22] J. Weston, A. Bordes, S. Chopra, A. M. Rush, B. van Merriënboer, A. Joulin, and T. Mikolov. Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks. *arXiv:1502.05698*, feb 2015.

Appendix

A.1 A note on general AI learning environments

Unlike infrequently published roadmaps and documents akin to our framework, increasingly more general environments have been recently introduced [2, 9, 4, 12, 18, 22, 13, 20, 8]. These vary in a number of ways, from the level of detail, dominant modality, to the complexity of the world within, each environment presents different possibilities and challenges.

In this short note, serving as a brief sample of the type of analyses the AI Roadmap Institute will perform on a daily basis, here we present a quick comparison of our realization of the agent development roadmap, our “School for AI” [20] with the recently introduced CommAI-env environment [13]:

The “School for AI” (SAI) environment [20] is a realization of a curriculum roadmap [17]. It provides a number of possible worlds, with different levels of complexity, continuity and purpose. The environment allows for defining learning curricula that encourage gradual and guided learning. First, a set of learning tasks, also called a “curriculum”, is designed. The aim of the curriculum is to teach the agent useful skills and abilities, so it does not have to discover them on its own. When the curriculum is ready, an agent is subjected to training. The agent’s performance on learning tasks in the curriculum is then evaluated and is used to improve the curriculum itself as well as the agent’s architecture.

The recently released CommAI-env (CAI) environment [13] shares with SAI the goal of supporting learning for general intelligence. The approaches are similar also in their emphasis on gradual learning with curricula of increasingly complex tasks, where the first tasks are often very simple for humans. The environments however differ in the information they make available to the agent. In SAI, the agent is immersed in a visual world, whereas in the CAI environment, it only has access to messages from the teacher. The approach is characterized in [13] as “language centric”. This suggests that the SAI supports more directly the learning of subsymbolic representations and sensory grounding of language. The question of how much sensory grounding is required for language understanding has yet to be settled, so CAI and SAI are complementary in a very interesting way.

Central to curriculum-based approaches to general AI are the problems of generating tasks, assessing their complexity, and determining the order between them. Efforts towards addressing these challenges are described in [1, 16, 21, 7]. To the best of our knowledge, these problems have yet to be discussed in the context of the CAI environment.

We suspect that the language-centric approach can eventually make more difficult the problems of task comparison and complexity measurement. In a perception-based approach, there is a natural progression of increasingly complex pattern recognition tasks. Once the agent starts learning language, the way it models meaning is biased by its history of living in a physical world. In a language-centric approach, there may be fewer constraints on how the agent can represent the world. While this may become the source of surprising insights, it can also make it harder to assess if a task is easy or hard.