Chapter 1
Introduction

Theories may be equivalent in all their predictions and are hence scientifically indistinguishable. However, different views suggest different kinds of modifications which might be made and hence are not equivalent with respect to the hypotheses one generates from them.

Richard P. Feynman, Nobel Lecture 1965

The rise of civilization is synonymous with the creation of tools that extend the intellectual and physical reach of human beings [133]. The pinnacle of such endeavours is to replicate the flexible reasoning capacity of human intelligence within a machine, making it capable of performing useful work on command, despite the complexity and adversity of the real world. In order to achieve such Artificial Intelligence (AI), a new approach is required: traditional symbolic AI has long been known to be too rigid to model complex and noisy phenomena and the sample-driven approach of Deep Learning cannot scale to the long-tailed distributions of the real world.

In this book, we describe a new approach for building a situated system that reflects upon its own reasoning and is capable of making decisions in light of its limited knowledge and resources. This reflective reasoning process addresses the vital safety issues that inevitably accompany open-ended reasoning: the system must perform its mission within a specifiable operational envelope.

We take a perspective centered on the requirements of *real-world* AI, in order to determine how well mainstream techniques fit these requirements, and propose alternative techniques that we claim have a better fit. To reiterate: by AI we mean the property of a machine that exhibits general-purpose intelligence of the kind exhibited by humans, i.e., enjoying the ability to continually adapt existing knowledge to different domains. The endeavor to create intelligent machines was definitively proposed as such in the 1950s [220], although the concept of a humanoid automaton recurs throughout recorded history. Due to the sheer magnitude and ambition of the project, there have naturally been many bumps in the road: not only the infamous 'AI winter' [202], but also periods where the endeavor's vision and direction have been clouded by the prospects of short-term success.

© The Author(s) 2022

J. Swan et al., *The Road to General Intelligence*,
Studies in Computational Intelligence 1049,
https://doi.org/10.1007/978-3-031-08020-3_1

AI for Automation

Given that substantial resources are required to create AI, it cannot be done on a whim. Therefore the shape of AI (at least in its initial incarnation) will be strongly influenced by the return anticipated by those investing in it. That is, to answer "How to build AI?", we must ask why we want AI in the first place, i.e., what is the *business case* for a machine with general intelligence?

Philosophical considerations aside, intelligent machines are ultimately tools for implementing a new leap in *automation*. In practical automation settings, the generality of a system is measured as the inverse of the cost of its deployment and maintenance in a given environment/task space. At the low end of this spectrum are systems that depend on full specifications of their environments and tasks. Such systems are very costly to re-deploy when facing specification changes, possibly incurring the highest cost: that of a complete rewrite. At the high end are more general systems that re-deploy autonomously through continual open-ended adaptation and anticipation.

The main functional requirement of general intelligence is therefore to *control the process of adaptation*. In this work, we claim that this can be achieved in a unified, domain-agnostic manner via the ability to ground arbitrary symbols (whether arising from end-user vocabulary or being synthesized by the system) in an explicit learned semantics. Hence, throughout this work, when we discuss *symbols* in reference to our proposed architecture, it is not in the sense of the a priori opaque logical predicates of 'Good Old-Fashioned AI', but rather follows in the footsteps of a collection of cyberneticists, psychologists and systems theorists [8, 67, 218, 253, 261, 269, 299] for whom *"symbols are merely shorthand notation for elements of behavioral control strategies."* [49].

In practical terms, the endeavor of creating general intelligence therefore consists of building a template for a *learning control system* which can be re-targeted at an arbitrary environment, bootstrapping the control mechanisms with as little latency as possible, starting from small amounts of (incomplete or even faulty) knowledge. The system is then expected to discover further constraints on the fly—be it from a corpus of ready-made knowledge; from experience acquired with and without supervision; perhaps by interacting in the environment, possibly under the sporadic guidance of teachers and end-users.

Notwithstanding these business considerations, the creation of AI still relies on good science, especially with regards to requirements engineering, with the initial focus illustrated in Fig. 1.1. Although we set aside those requirements that are mostly issues of hardware, paperwork, or procedures (e.g., constructing curricula for teaching the system as well as its eventual operators), the fact that they must be addressed and fulfilled then imposes constraints on which scientific techniques can even be considered. The requirement-centric perspective dictates which properties are important for a technique to exhibit or avoid. For example, even legal requirements impinge on techniques, such as when GDPR[1] demands transparency in automated decision

[1] General Data Protection Regulation (EU) 2016/679.

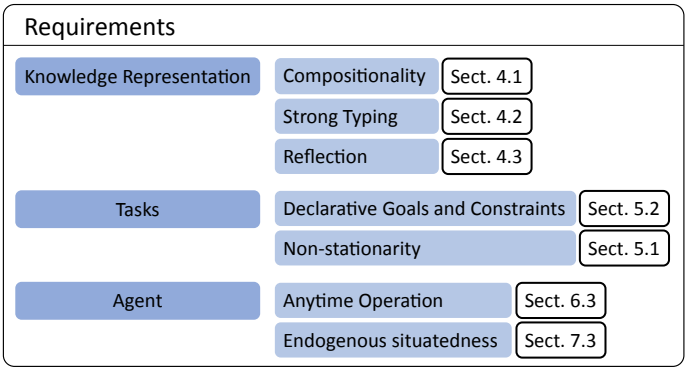

Fig. 1.1 Theme development in Part I: a summary of the most pertinent engineering requirements for constructing a general intelligence system of real-world use. Their importance is established throughout the first part of this book and then leveraged to construct our proposed framework: Semantically Closed Learning

making, which is more easily fulfilled when knowledge representation and reasoning are not intrinsically black-box components or processes.

The Structure of this Book

This book begins with a survey of historical (Chap. 2) and contemporary (Chap. 3) AI methodologies, discussing their strengths and weaknesses, from the perspective of their potential to support general intelligence. Machine learning (ML), notably deep- and reinforcement learning, has emerged as the dominant AI paradigm. There are certainly many valuable applications for which ML offers functionally good solutions, in particular for industrial applications where such techniques are used to build control systems beyond the reach of traditional software engineering. Nevertheless, it remains a feat of imagination to ascribe any meaningful notion of intelligence to any of these systems: the constraints and ambitions of machine learning and general intelligence research are simply orthogonal. Although machine learning is a valuable *engineering technique*, this fact is not to be confused with a *claim* that it might offer a path toward general intelligence. In Chaps. 4 and 5, we make a critical appraisal of this claim, by contrasting deep learning and reinforcement learning techniques against key requirements of general intelligence—from the perspective of automation engineering, these are reified by the notion of 'Work on Command' in Chap. 6.

The second part of the book is concerned with an alternative framework that we claim fulfills better these requirements. There is increasing consensus that it is necessary to combine the strengths of both symbolic and connectionist paradigms [59, 210]: the main advantage of symbolic approaches is the ready injection of domain knowledge, with the attendant pruning of hypothesis space. In contrast, the main advantage of connectionism is that it is (at least in principle) a *tabula rasa*.

As has been argued by Marcus for many years [214], we also hold the view that general intelligence requires the recursively algebraic capacities of human reasoning. This motivated the research and associated reference architecture implementation we present in this book. This architecture has been implemented, and prototypes have been developed, addressing the domains of medical diagnosis, service robotics, and industrial process automation—empirical demonstrations will be the topic of subsequent works. In Chaps. 7–10, we define a framework for 'Semantically Closed Learning' which:

- Describes an explicit (but nonetheless 'universal') recursive interpreter for a highly generalized notion of algebraic reasoning.
- Represents the hierarchical causal structure of hypotheses as first-class objects.
- Defines a fine-grained and resource-aware attention mechanism, driven to favor highly-structured and stable hypotheses.
- Describes key reasoning heuristics using the generic and compositional vocabulary of category theory, from the emerging perspective of 'Categorical Cybernetics' [42, 140].
- Defines a novel compositional mechanism, using *lenses* [88], an approach which unifies conventional backpropagation, variational inference, and dynamic programming, for the purpose of abductive reasoning over hybrid numeric-symbolic expressions.
- Describes a minimal viable implementation design for 2nd *order automation engineering*—system identification, synthesis, and maintenance—with *guarantees* relevant to safety.

Finally, in Chap. 11, we summarize our contribution, discuss research avenues, and conclude.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.