Artificial Intelligence 221 (2015) 36–45

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Wanted: Collaborative intelligence

Susan L. Epstein [a,b,*]

[a] Department of Computer Science, Hunter College, New York, NY 10065, USA
[b] The Graduate Center of The City University of New York, 695 Park Avenue, New York, NY 10016, USA

ARTICLE INFO

Article history:
Received 5 May 2014
Received in revised form 20 November 2014
Accepted 17 December 2014
Available online 6 January 2015

Keywords:
Collaboration
Intelligence
Perception
Dialogue

ABSTRACT

Although the original vision for artificial intelligence was the simulation of (implicitly human) intelligence, research has gradually shifted to autonomous systems that compete with people. The resultant popular attitude toward artificial intelligence, we argue here, is by turns disdain, grudging acceptance, and fear. That attitude not only limits our work's potential, but also imperils its support. This paper proposes a constructive alternative: the development of collaborative intelligence. As envisioned here, a collaborative intelligence does not require encyclopedic command and need not be limited to a single problem. The necessary components of a collaborative intelligence are nearly at hand, and the key issues readily identified. As a first step, this paper proposes three challenging but accessible problems that would both change the public perception of artificial intelligence and spur substantive research to advance our science.

1. Introduction

This is a position paper, intended both to challenge and to inspire research in artificial intelligence (*AI*). It considers the path AI research has taken, and seeks to broaden its long-term goals. The vision offered here integrates what AI has learned in a way that would advance our science and support people.

This section restates the core ideas that began AI, describes how we as researchers now see our field, considers the current media portrayal of AI, and notes how our research targets shift. In response, Section 2 proposes collaboration between a person and a machine, and highlights some crucial differences between human and computer collaborators. An extended example of collaborative intelligence in Section 3 identifies key issues inherent in the development of a collaborative intelligence. Section 4 presents three reasonably approachable, exciting problems that would both engage AI researchers and benefit human users.

1.1. How AI was first envisioned

In August 1955, in what has come to be known as the Dartmouth manifesto, Claude Shannon, Marvin Minsky, Nathaniel Rochester, and John McCarthy proposed AI as the theme for a conference to be held the following year.[1] Their targets for this new field included precursors of what AI would now term problem solving, natural language processing, artificial neural networks, complexity theory, machine learning, and perception. Their premise was that "every aspect of learning or

* Corresponding author at: Department of Computer Science, Hunter College, 695 Park Avenue, New York, NY 10065, USA. Tel.: +1 212 7725213.
E-mail address: susan.epstein@hunter.cuny.edu.

[1] http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, May 1, 2014.

http://dx.doi.org/10.1016/j.artint.2014.12.006
0004-3702/

any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." There is no indication in that document that computers were to work together with people; the intent was to develop an autonomous intelligence.

At the time, Alan Newell and Herbert Simon had already begun work on the Logic Theorist, a program to prove mathematical theorems. Although it could prove 38 of the first 52 theorems in *Principia Mathematica* (a key contribution to the foundations of mathematics [31]), the Logic Theorist was of little interest at the Dartmouth conference [7]. Few scholars then considered computer-constructed proofs of elementary theorems a dramatic achievement. Nonetheless, Newell and Simon, along with J.C. Shaw, continued their work. They were quite clear about the Logic Theorist's need for knowledge, both about the problem domain and the mechanisms necessary to guide its search. This led to the General Problem Solver (*GPS*) [20]. GPS tackled problem solving from two perspectives: machine-achievable processes and behavior observed in humans. True to the Dartmouth vision of simulation, college sophomores were asked to think aloud as they solved symbolic logic problems, and GPS was developed to simulate what Newell and Simon observed.

In many ways, in 1958 GPS set the agenda for AI research. It established some of AI's basic building blocks: goals, objects characterized by feature values, and operators as functions on objects. GPS explicitly recognized the importance of feature identification, problem-dependent heuristics, efficient data structures and algorithms, performance evaluation, and the potential for exponential search spaces. In response, it advocated problem-independent heuristics, including means-ends analysis and planning; a model of problem solving as a sequence of actions; and the careful separation of process from data (i.e., isolation of *what* was being thought about from *how* it was being thought about). We consider next how AI has addressed that agenda.

1.2. AI's trajectory

The path AI research has taken can be readily traced from the first AI conference at Dartmouth through two of the major meetings that followed it. The Dartmouth manifesto postulated general goals: to discover "how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves." There was no mention of interaction with or impact on people, and no targets beyond the things that people can do.

Fourteen years later, the papers presented at the first IJCAI in 1969 reflect the topics AI researchers chose to pursue. The Wordle™ in Fig. 1(a) for the titles of those papers indicates a strong focus on implementation (e.g., the tokens "system," "program," "memori," "method," "process," "language," "robot," "structure," and "control") and theory (e.g., "model," "theorem," "base"). There is also a significant interest in "pattern," a precursor to learning.

Contemporary AI is described by a second Wordle, in Fig. 1(b), this time for the titles of papers presented at the 23rd IJCAI, in 2013. "Learn" and "model" now clearly dominate, with techniques (e.g., "plan," "gener," "search," "constraint," "logic," and "network") as important concerns, along with "game," "social," and "data." This is how we, as researchers, now describe our field to one another. Several subfields, most notably vision and robotics, have by 2013 detached themselves, but a focus on language and structure persists. Although "human" appears in both Wordles, it is not prominent in either of them.

The dramatic changes between 1969 (in Fig. 1(a)) and 2013 (Fig. 1(b)) were driven, I believe, by our collective fascination with hard problems. To solve those problems, AI researchers developed a diverse set of representations to model the real world for computers. These include logics, ontologies, semantic nets, and rich graph structures. Then, to harness these representations, AI researchers built inference mechanisms and search algorithms intended to manipulate that knowledge. Empirical work rigorously validates the efficiency and effectiveness of these products.

The combined push to compare and compete has honed performance and driven research. The need for uniform scientific evaluation of empirical work has led to ambitious, shared datasets, including repositories for machine learning, planning, and image labeling. Competition, along with common and exacting evaluation metrics, allows us to see which methods perform best on which data. The earliest instance of this appears to have been the push to prove all the theorems in *Principia Mathematica*, a feat ultimately accomplished independently by both Gilmore and Wang [9,30]. Since then, prize money and targeted research funding have engendered enthusiastic competitions in such areas as speech recognition, game playing, and self-driving cars, where work might have otherwise developed far more slowly.

As a result, AI's standard for success has become the ability of one system, algorithm, architecture, representation, or approach to outperform another. Clearly, we are in search of the best machine intelligence we can construct, without any regard to what people can do. Meanwhile, this clever problem solving has had some unanticipated results.

1.3. AI in the mainstream media

AI's description in the popular press has done little to enhance its reputation. The media has alternately overpraised AI for techniques that are not new, and over-criticized it for overly optimistic promises.[2] As LeCun notes, "AI 'died' about four

[2] Compare, for example, http://www.dailygalaxy.com/my_weblog/2010/01/artificial-intelligence-will-leapfrog-humans-by-2020-says-scifi-great-a-weekend-feature.html with http://www.skeptic.com/reading_room/artificial-intelligence-gone-awry/, January 7, 2015.