While researchers in AI all strive to create intelligent machines, separate AI communities view intelligence in strikingly different ways. Some abstract intelligence through the lens of connectionist neural networks, while others use mathematical models of decision processes or view intelligence as symbol manipulation. Similarly, researchers focus on different processes for generating intelligence, such as learning through reinforcement, natural evolution, logical inference, and statistics. The result is a panoply of approaches and subfields.

Because of independent vocabularies, internalized assumptions, and separate meetings, AI subcommunities can become increasingly insulated from one another even as they pursue the same ultimate goal. Further deepening the separation, researchers may view other approaches only in caricature, unintentionally simplifying the motivations and research of other researchers. Such isolation can frustrate timely dissemination of useful insights, leading to wasted effort and unnecessary rediscovery.

To address such dangers, we organized an AAAI Fall Symposium called “How Should Intelligence Be Abstracted in AI Research” that gathered experts with diverse perspectives on biological and synthetic intelligence. The hope was that such a meeting might lead to a productive examination of the value and promise of different approaches, and perhaps even inspire syntheses that cross traditional boundaries. However, organizing a cross-disciplinary symposium has risks as well. Discussion could have focused narrowly on intractable disagreements, or on which singular abstraction is “the best.” An unhelpful slugfest of ideas could have emerged instead of collaborative cross-pollination, leading to a veritable AI Tower of Babel.

In the end, there were world-class keynote speakers spanning AI and biology (see Table 1), and participants were indeed collaborative. Some traveled to the United States from as far as Brazil, Australia, and Singapore; but beyond geographic diversity, there were representatives from many disciplines and approaches to AI (see Figure 1). Drawing from the symposium’s talks and events, we now summarize recent progress across AI fields, as well as the key ideas, debates, and challenges identified by the attendees. (See also the sidebar, “Straight from the Experts,” which showcases and summarizes the direct viewpoints of some of the keynote speakers.)

**Key Ideas Discussed**

One controversial topic was deep learning, which has recently shattered many performance records over an impressive spectrum of machine learning tasks. The central idea behind deep learning is that large hierarchical artificial neural networks (ANNs), inspired by those found in the neocortex, can be trained on big data (for example, millions of images) to learn a hierarchy of increasingly abstract features (see Figure 2). Overall, participants agreed that recent progress in deep networks was a significant step forward for processing streams of high-dimensional raw data into meaningful abstract representations, which is required for tasks like recognizing faces from unprocessed pixel data. But there was also agreement that much work remains to create algorithms that leverage such representations to produce intelligent behavior and learn in real-time from feedback; in other words, scaling deep learning to more cognitive behavior may prove problematic.
Andrew Ng, affiliated with Stanford University and Baidu Research, gave a keynote on deep learning that outlined its motivation, implementation, and recent successes. Other keynote speakers reported that they also effectively use deep learning, in that their research similarly involves learning in many-layered neural networks. In this sense, deep learning has gone by many names over time, and is currently being reinvigorated by increased computing power, Big Data, greater biological understanding, and algorithmic advances. For example, in his keynote, Randall O’Reilly of the University of Colorado at Boulder summarized his work in the field of computational neuroscience, where researchers often develop cognitive architectures, which are computational processes designed to model human or animal intelligence. His Leabra cognitive architecture is a many-layered neural network modeled on the human brain, which includes collections of neurons analogous to the major known functional areas of the brain. In this way, two separate areas of AI apply similar technologies inspired by different motivations: one coarsely abstracts brains to solve practical problems, and the other applies more biologically plausible abstractions to better understand animal brains.

A related camp (to which the authors belong) that’s inspired by nature and applies evolutionary algorithms to design neural networks, is called neuroevolution. In his keynote, Risto Miikkulainen of the University of Texas at Austin described how neuroevolution can design cognitive architectures via a bottom-up design process guided by evolutionary algorithms instead of through top-down human engineering. Kenneth Stanley, from the University of Central Florida, argued that evolutionary approaches may be important tools for producing human-level AI because evolution is highly adept at creating variations on an underlying theme. The idea is that evolutionary methods could perhaps provide this important capability to other AI techniques, such as deep learning. Supporting this idea, Jeff Clune, from the University of Wyoming, described how evolutionary algorithms that incorporate realistic constraints on natural evolution can produce ANNs that have important properties of complex biological brains, like regularity, modularity, and hierarchy.

Pierre-Yves Oudeyer of Inria detailed in his keynote the field of developmental robotics, which investigates how robots can develop their behaviors over time through interacting with the world, just as animals and humans do. Representative approaches in developmental robotics implement mechanisms to enable lifelong, active, and incremental acquisition of both skills and models of the environment, through self-exploration or social guidance. Oudeyer’s research shows that motivating robots to be curious results in continual experimentation: A robot equipped with intrinsic motivation will search for information gain for its own sake; at any given point in the robot’s development, it actively performs experiments to learn how its actions affect the environment. Because such curiosity leads to an ever-improving model of the consequences of a robot’s actions, over time it can result in learning how to accomplish increasingly complex tasks.

Among the traditional biologists that attended was Georg Striedter of the University of California at Irvine, author of the influential book Principles.
The following keynote speakers weighed in with their diverse perspectives on biological and synthetic intelligence.

**Pierre-Yves Oudeyer, Inria and Ensta ParisTech**

Artificial intelligence has been struggling with two major mistakes. First, it has conflated human capabilities to think, feel, and act with a context-independent concept of “general intelligence.” This is wrong from a biological and psychological point of view: like all animals, humans are equipped with cognitive mechanisms that are highly adapted to the families of changing environments in which they live. These mechanisms are powerful, but in no way “general”: we’re skilled at what we need in our ecosystem (such as interpreting social behavior), but poor at other things (such as numerically solving differential equations). Learning theory also tells us that general intelligence doesn’t exist: solving difficult problems with limited time resources requires biases.

A second mistake is that researchers have been focusing on particular information-processing techniques at single levels of abstraction. But we know that even the non-general intelligence of humans cannot be understood through reductionist ethereal approaches. Sensorimotor, cognitive, and social capabilities in the child self-organize out of dynamic interactions within and across the brain, the body, and the physical and social environment, and over multiple spatiotemporal scales. Adaptive thinking and acting is an embodied, situated, and dynamic complex system.

Thus, identifying a precise target ecosystem and context of operation should be crucial to any attempt to build advanced cognitive machines. One possibility is to target the human-like capabilities in the human ecosystem, and to attempt modeling the interaction of multiple mechanisms (for example, maturation, motivation, learning, physical dynamics, and reasoning) at different scales of time and abstraction to guide the progressive development of certain families of skills (such as co-development of language and action in a social context). This is what fuels the emerging fields of evolutionary and developmental robotics (see http://en.wikipedia.org/wiki/Developmental_robotics).

**Risto Miikkulainen, University of Texas at Austin**

Intelligent behaviors and neural systems that generate them didn’t emerge in a vacuum. They resulted from evolution and development in complex environments where they were embodied in physical structures, interacted with other behaviors, and were continuously changing. To understand biological intelligence, it is thus necessary to take into account how it emerged over time—that is, how key evolutionary stepping stones and adaptive pressures determined what we see today. In order to build artificial intelligence systems that rival biology, it’s useful to follow the same path. With today’s computational power, we can create complex embodied, changing, multiagent environments and study how cognitive architectures emerge in them. The challenge is thus not how intelligence should be abstracted, but how the environment should—intelligence will then follow.

**Randall O’Reilly, University of Colorado at Boulder**

Here’s a provocative claim: the computer science (CS) approach to AI tends to be much more trend-driven than the cognitive neuroscience (CN) approach: CS folks tend to swarm around the latest best-performing algorithm. Hence, there’s the current fascination with deep networks (and support-vector-machines before them, and so on). In contrast, CN folks are more swayed by theoretical constructs that integrate large quantities of data, and these tend to be more slowly evolving and admitting of a greater plurality. For example, we connect strongly with the ACT-R folks around a shared view of the central role of the basal ganglia in orchestrating the flow of cognition, but their model lacks a proper hippocampus of the sort that we connect strongly with other researchers around. We in CN also don’t believe that there’s just one killer algorithm at the heart of cognition: there are many, working together in complex, spatiotemporal scales. Adaptive thinking and acting is an embodied, situated, and dynamic complex system.

Thus, identifying a precise target ecosystem and context of operation should be crucial to any attempt to build advanced cognitive machines. One possibility is to target the human-like capabilities in the human ecosystem, and to attempt modeling the interaction of multiple mechanisms (for example, maturation, motivation, learning, physical dynamics, and reasoning) at different scales of time and abstraction to guide the progressive development of certain families of skills (such as co-development of language and action in a social context). This is what fuels the emerging fields of evolutionary and developmental robotics (see http://en.wikipedia.org/wiki/Developmental_robotics).

These days, there’s much enthusiasm in AI and most of it comes from machine learning; techniques like Deep Learning have, with the aid of GPUs and Big Data, become a source of big profits and record-setting results, in domains such as speech recognition and image recognition.

But another Winter could easily come. In core domains like reasoning and natural language understanding, there has been less progress; perhaps nothing notable since the impressive, but still limited, Watson and Siri. Machines still can’t match toddlers at acquiring language. General-purpose robots like the fictional Rosie still seem like a very long way away, even with advances in machine learning.

Part of the problem, in my view, is that machine learning itself is cast too narrowly; most efforts focus on improving techniques for classification, determining which of some previously known set of categories a particular example (say, a handwritten digit) belongs to. But human reasoners routinely go beyond what they have seen before, drawing inferences that have never been made, and producing sentences that have never been said; human cognition routinely extends finite mechanisms to infinite possibility.

The only viable account of this starts with the notion of an abstract algebra of generalization—that we learn abstract rules that we can extend to arbitrary instances of variables. How do we do this, is, frankly, a mystery.

Until we unravel that mystery, what passes for learning in AI will remain too weak; machines will likely remain as savants, skilled at narrow tasks, but with no genuine understanding of language or the world.
of Brain Evolution (Sinauer Associates, 2005). His keynote focused on the history of how brain functionality has been viewed over time. He noted an interesting parallel between the history of AI and of neuroscience: In both, a simple serial view of intelligence led to exploring more parallel, distributed notions of processing. He mentioned that Rodney Brooks’ subsumption architecture in particular had influenced him, because it offered a picture of higher-order thought beyond simplistic linear pathways; while computer scientists often debate the promise of various approaches to computational intelligence among themselves, it’s informative also to consider the opinions of those who study how it arose in humans.

Aside from models with concrete biological inspiration, other attendees focused on abstractions of intelligence based on Markov decision processes (MDPs) and less-restrictive generalizations called partially observable Markov decision processes (POMDPs). Such MDPs and POMDPs represent decision making in a mathematical framework composed of mappings between states, actions, and rewards. This framework provides the basis for AI techniques like reinforcement learning and probabilistic graphical models. Devin Grady of Rice University and Shiqi Zhang of Texas Tech University each described mechanisms to augment such techniques to allow them to better scale to more complex problems. A similar need for tractable models motivated Andrew Ng’s change in focus from MDP-based reinforcement learning to deep learning. He mentioned in response to a question that he felt the bottleneck was no longer reinforcement learning algorithms themselves, but in generating strong relevant features from raw input for such algorithms to learn from, which otherwise must be manually generated by humans through domain-relevant knowledge.

Proponents of symbolic AI (also known as GOFAI, or “good old-fashioned artificial intelligence,” due to its early research dominance) defended their view that the power of human intelligence is largely captured in the idea of symbol manipulation. Such researchers also illuminated where non-symbolic approaches still fall short. In particular, John Laird of the University of Michigan posed an interesting challenge problem called embodied taskability: Similar to learning from demonstration, a robot must learn to perform novel tasks by interacting with humans. The task is intriguing because it’s an ambitious problem not often tackled by other fields of AI, yet it’s characteristic of human intelligence. Complementarily, Gary Marcus of New York University gave a provocative keynote highlighting several capabilities necessary for strong AI that current high-performing connectionist approaches do not yet implement, such as representing causal relationships and abstract ideas, and making logical inferences. He also mentioned challenges in natural language understanding.

During one of the panel sessions, an idea was proposed in an attempt to tie all of these fields and levels of abstraction together: a stack of models, where each individual level of the stack is guided by a different level of abstraction. The idea is that with such a stack the various levels of abstraction could be linked together, guided by a reductionist goal of connecting understanding of high-level, abstract, rational components of intelligence to “lower-level” ones that are closer to perceiving raw data and controlling muscles. For example, high-level
GOFAI algorithms could possibly be connected to deep learning models, which could be connected to more biologically plausible computational models of brains. In this way, it might be possible to unite disparate views and approaches to gain greater overall understanding.

**Debates**

As mentioned previously, deep learning proved to be a lightning rod for discussion and many researchers were quick to point out perceived difficulties in scaling deep learning to human-level AI. Open research questions include how to create deep networks that implement reinforcement learning, develop higher cognitive abilities over time, or manipulate symbols. Andrew Ng, when asked about merging deep learning with reinforcement learning, responded that it’s an unsolved problem and that “a seminal paper on that subject is waiting to be written.” (We should note here that the symposium occurred in November 2013, and since then a paper has gained significant attention that combines deep learning and reinforcement learning.\(^\text{11}\)) Ng was hopeful that it should be possible to extend deep learning algorithms to perform reinforcement learning without merging in other AI paradigms. In contrast, and perhaps unsurprisingly, researchers outside of deep learning were generally more skeptical.

While the current winds of AI seem generally to favor statistical machine learning methods like deep learning or reinforcement learning over purely symbolic GOFAI approaches, proponents of symbolic AI made convincing arguments for its continued relevance. John Laird expressed that although symbolic AI might not be as dominant as it once was, research progresses onward irrespective of current fashion. In particular, symbolic AI research currently is producing promising symbolic cognitive architectures that can empower agents to learn new human-taught tasks. In his keynote, Gary Marcus argued that it would be a mistake to conflate the time of an approach’s first prominence with its potential; he noted that symbolic AI techniques might also (like statistical techniques) benefit from advances in computing power and available data, and that such symbolic techniques were developed mainly in the absence of the broad computational resources that are now used in statistical approaches.

A point of agreement was that symbolic AI isn’t better or worse than alternate approaches, but is instead different in its aims and objectives. Symbolic AI continues to aim at the ambitious goal of general artificial intelligence (that is, human-level intelligence) while other approaches often focus on narrower domains or simpler forms of intelligence. A contribution of Gary Marcus was to highlight that GOFAI isn’t an inferior way of reproducing these narrower or simpler intelligences, but is instead aimed at a different goal: the cognitive intelligence that sets humans apart from other animals, which is where statistical machine learning methods are arguably weakest.

A contentious issue for researchers in biologically inspired AI concerned which biological details are extraneous and therefore unnecessary to include in AI models. For example, brains vary over a multitude of dimensions including neuron size, density, type, connectivity, and structure; intuitively, it seems unlikely that all such dimensions are equally important to a model’s functionality. Randall O’Reilly mentioned that in his models, the additional complexity of simulating neurons with binary spikes in time (like biological neurons) provided little benefit over simpler neuron models. Yet in contrast, Oliver Coleman (University of New South Wales) highlighted past research showing that the timing of spikes may be an important facet of learning processes in the brain. Taking the attendees as a whole, there were more skeptics of complex neuron models than proponents, likely reflecting a cautious, pragmatic preference for simplicity over biological realism for its own sake.

The opposite question was also debated: Are there salient features of brains and intelligence that are unfairly ignored? For example, O’Reilly believes that glial cells, which are non-neural cells that provide support and protection for neurons, may be more important computationally than their absence in most models would suggest. For Risto Miikkulainen and Pierre-Yves Oudeyer, how brains physically develop over time was a topic deserving greater attention; most models ignore the fact that biological brains learn while they grow and develop into their full mature size. In contrast, Gary Marcus argued that it may be possible to abstract nearly all biological detail away if all we care about is engineering AI, and not understanding biology. The resulting discussion questioned whether the brain is a well-engineered machine with much to teach us, or whether it’s merely a hacked-together “kluge”.\(^\text{12}\) In other words, do researchers mistakenly idealize the human brain, searching for elegant insights in a messily designed artifact—one that’s functional but ultimately unintelligible?

As the debate became more intense, Pierre-Yves Oudeyer interjected that, of course, which biological details are important depends upon the scientific question being investigated. Or, as John Laird said in response to the name of the symposium (“How Should Intelligence Be Abstracted in AI Research?”), “It depends!” Oudeyer
then said something that resonated strongly: Because we don’t deeply understand intelligence or know how to produce general AI, rather than cutting off any avenues of exploration, to truly make progress we should embrace AI’s “anarchy of methods.”

**Major Challenges**

Through the course of the discussion, many remaining challenges for AI became evident that cut across traditional boundaries. Overall, AI approaches tend to have four distinct focuses: Real-world embodiment, building features from raw perception, making decisions based on features, and high-level cognitive reasoning that’s unique to humans. Approaches generally specialize on one such area, and often perform poorly when stretched beyond that focus. However, general AI requires spanning such divides. To do so may require integrating existing disparate technologies together; for example, hybrid neural systems often combine neural network and symbolic models together, like the SAL architecture that connects the symbolic ACT-R model to bottom-up perception from the Leabra neural model. A more conventional approach is to attempt to scale up an existing technology beyond its current borders. For example, Risto Miikkulainen’s keynote highlighted that neuroevolution techniques are beginning to evolve instances of simple cognitive architectures. Additionally, cognitive architectures like Leabra and Spaun are beginning to tackle symbolic manipulation of variables through human-engineered neural mechanisms. Extensions to deep learning might similarly incorporate decision making and cognition. However, if integrating or extending existing technologies proves unproductive, there might yet be a need for new approaches better able to bridge aspects of AI ranging from low-level perception to human-level cognition.

An interesting challenge in AI that often goes unconsidered is safety. The most interesting intellectual challenge drawing researchers to AI is understanding and engineering intelligent systems. However, it may be dangerous to single-mindedly pursue such a goal without considering the transformative consequences that may result if we create AI that rivals or even surpasses human intelligence. Problematically, academic and industrial incentives are nearly unilaterally aligned towards creating increasingly sophisticated AI, discounting through omission potentially important critical reflection on its dangers and unintended side effects. Only a single talk, by Armando Tacchella of the University of Genova, focused on creating safe abstractions of AI. That work raised difficult questions for the many AI approaches where verification or automatic characterization of the behaviors produced is difficult. For example, neural networks are notorious for being black box models, making interpreting the safety of agents resulting from deep learning, neuroevolution, and neural-based cognitive architectures difficult. A consensus among attendees was that this was an important and underfunded consideration.

Another central problem that emerged through discussions is the difficulty (or impossibility) of definitively knowing what ways of abstracting intelligence are truly “better” or more productive than others. In general, attempting to predict the future promise of any particular technology or research direction is often misleading. But a particular challenge in AI stems from the existence of only one example of high-level intelligence from which to infer generalities. As a result of nature’s singular anecdote on intelligence, separating what is essential for intelligence from what is merely coincidental remains difficult.

At the symposium’s end, researchers mentioned that they better understood the philosophical and theoretical motivations for areas of AI they had unintentionally only seen previously in caricature. One participant said that he learned that even when viewing intelligence abstractly from a high level, there’s a benefit to following key developments at lower levels. Another offered that he “learned how limited our knowledge is,” and that it was interesting how often “key leaders in a field might not have a grand, deep plan [...] but that instead, behind the curtain, are scientists doing the best they can, fumbling in the dark.” Another made reference to theparable of three blind men describing an elephant, where each blind man describes the whole elephant in terms of features specific to the individual parts they’re examining (a tail, a tusk, or a leg, respectively), which leads to very different interpretations of what an elephant is. Similarly, through sharing local perspectives on AI and what they imply about the overall field, the resulting traces of intelligence’s outline—made from all angles and levels of abstraction of AI’s anarchy of methods—might potentially be combined to accelerate our understanding of the general principles underlying intelligence and how to recreate it computationally.

**References**


---

**Joel Lehman** is a postdoctoral fellow at the University of Texas at Austin. He is an inventor of the novelty search algorithm. Other research interests include neuroevolution, artificial intelligence, robots, and physical designs. Lehman has a PhD in computer science from Michigan State University. Articles about his research have appeared in many news publications, including National Geographic, NPR, NBC News, Discover, the BBC, the New Scientist, The Daily Telegraph, Slashdot, MIT’s Technology Review, and U.S. News & World Report. More information about his research is available at http://joellehman.com.

**Jeff Clune** is an assistant professor in the Computer Science Department at the University of Wyoming, where he directs the Evolving AI Lab. He studies evolutionary computation, a technology that harnesses natural selection to evolve, instead of engineer, artificial intelligence, robots, and physical designs. Clune has a PhD in computer science from Michigan State University. Articles about his research have appeared in many news publications, including National Geographic, NPR, NBC News, Discover, the BBC, the New Scientist, The Daily Telegraph, Slashdot, MIT’s Technology Review, and U.S. News & World Report. More information about his research is available at http://jeffclune.com.

**Sebastian Risi** is an assistant professor at the IT University of Copenhagen. His interests include neuroevolution, evolutionary robotics and design automation. Risi has a PhD in computer science from the University of Central Florida. He has won several best paper awards at GECCO and IJCNN for his work on adaptive systems and the HyperNEAT algorithm for evolving complex artificial neural networks. He’s also a co-founder of FinchBeak, a company that creates casual and educational social games enabled by next-generation AI technology. More information about his research can be found at http://sebastianrisi.com.