Giant Shift Forward

Jonathan Berant navigates a new paradigm in natural language processing and artificial intelligence

By Alex Hutchinson

Photographs by Boaz Perlstein

In the fall of 2019, Google tweaked its search algorithm. The company knew that people tended to type their queries in "keyword-ese," rather than phrasing them the way they would speak to another human, so its researchers developed a new technique that sought to glean meaning from whole phrases or sentences rather than individual keywords. With this change, when presented with a search like "brazil traveller to canada need visa," Google can now spot the crucial word "to," assess its context, and return only results about travelling from Brazil to Canada and not vice versa.

Such advances often feel small or incremental. Who has not become a bit blasé about the steadily growing competence of Siri and her virtual peers? But Google's development, which it dubbed Bidirectional Encoder Representations from Transformers, or BERT, marked a bigger step. "Everything changed afterwards," says Jonathan Berant, a former postdoctoral researcher at Google who is now a professor at Tel-Aviv University's Blavatnik School of Computer Science. "And BERT is the model that started this revolution."

Berant, who started his PhD at Tel Aviv University as an Azrieli Graduate Studies Fellow in 2007, the first cohort of the program, studies natural language processing, a field that has always loomed large in our conception of what it means for a machine to be "artificially intelligent." Most famously, the British mathematician Alan Turing proposed in 1950 that a computer's ability to carry on a human-like conversation would be a reasonable proxy for whether the computer could "think."

The field has undergone several dramatic shifts since Berant finished his doctorate: chatty digital assistants have become ubiquitous, the Turing Test was beaten (albeit controversially) in 2014, and the success of new approaches like BERT has forced everyone in natural language processing to rethink their research agendas. "When a field is exploding, you need to think more about what it is that you do exactly," Berant says. "What is your advantage?"

Berant's interest in natural language processing began with a linguistics course he took at the Open University during his military service. "My head exploded," he recalls. "I thought this was amazing." He wanted to pursue the topic, but he also recognized that he had some "exact science tendencies," so after his service was finished, he enrolled at Tel Aviv University in a joint computer science and linguistics program. Four years later, Berant started his PhD as an Azrieli Fellow, eventually zeroing in on a problem in natural language processing called textual entailment.

Given two statements, can you infer one from the other? To humans, it's clear that if Amazon *acquires* MGM Studios, that means that Amazon *owns* MGM Studios. But these leaps are trickier for a computer to make. After all, if you acquire a second language, that doesn't mean that you own it. Berant's thesis focused on using the underlying structures of language — properties like transitivity, which means that if A implies B and B implies C, A must imply C — to help computers make better inferences. Despite all the progress over the past decade, textual entailment is a problem that researchers are still grappling with, Berant says: "It kind of encapsulates a lot of the things that you need to do in order to understand language."

After completing his PhD, Berant headed to Stanford University in California for a postdoctoral fellowship. He began working with computer scientists Percy Liang and Christopher Manning and shifted his focus to semantic parsing, which is the task of taking a single sentence of natural language and translating it into a logical form that a computer can understand and act upon. If you tell the virtual assistant on your phone, "Book me a ticket on the next flight to New York, but only a morning flight, with no connections," that's a very specific set of instructions that the software has to understand, no matter how you phrase it.

The first attempts to build computer systems that could understand natural language, starting in the 1960s, relied on rules. In a sufficiently narrow domain, you could tell the computer everything it needed to know in order to answer questions. But that approach has limits. "There's a lot of world out there," as Liang once put it, "and it's messy too."

In the 1990s, these rules-based systems were supplanted by statistical approaches, in which computers were programmed to "learn" by adjusting their own parameters after, say, answering a question correctly or incorrectly. To outperform rules, the statistical approach requires huge numbers of human-generated examples. While at Stanford, Berant and his colleagues used Google Suggest to generate a million sample questions, then enlisted human workers on Amazon Mechanical Turk to answer 100,000 of them at three cents per question. This approach can be enormously powerful: it's how IBM's computer system Watson beat record-holding game show champion Ken Jennings on *Jeopardy* in 2011. But the statistical approach, too, is limited by the sheer scale of human-provided data required.

What makes BERT and its successors (known collectively by the proposed name "foundation models") so special is that they're easy to train. A 2012 breakthrough led by Google's Quoc Le showed that if you use a big enough neural network — a type of algorithm in which the flow of information is modelled after the nerves and synapses of the human brain — then you don't need to feed it specially prepared data with pre-labelled answers. Instead, you can simply feed it a massive pile of previously existing data, like ten million still frames lifted from YouTube videos, and the algorithm learns to recognize recurring features, such as cats, without ever being told what a cat is. This approach is known as unsupervised learning, and it was quickly adopted in natural language processing with great success by using existing troves of fact-rich natural language like Wikipedia. "Everything became neural networks in natural language processing," Berant says, "and it works very well for most things — much better than everything we had before."

In a word, BERT's impact was "huge," says Liang, who launched the Center for Research on Foundation Models at Stanford in August 2021. "Shortly after it came out, essentially all state-of-the-art natural language processing models became based on BERT or some other foundation model." There's now a Hebrew version of BERT, AlephBERT, and a French version, CamemBERT. The success of these models sparked a flurry of interest that crossed discipline boundaries. BERT's successors are being used in neighbouring fields such as computer vision and robotics, and for more exotic applications such as predicting the three-dimensional structure of proteins, where their performance is revolutionizing the field.

But the scale of computing resources required has created a dilemma for academics like Berant. The Google algorithm introduced in 2012 used a neural network with more than a billion synapses that ran on a cluster of 1,000 powerful computers yoked together into a single system. The price of entry also continues to climb: the current state-of-the-art neural network, known as GPT-3, boasts a staggering 175 billion synapses. "The places where you can actually build models that are the best in the world are now more or less restricted to Google, Facebook and Microsoft," Berant says. "So, academia needs to reposition itself and figure out: What is our role?"

Who has not become a bit blasé about the steadily growing competence of Siri and her virtual peers?

Jonathan Berant is working on the technical challenges of natural language processing, such as reasoning in computer programs, while pondering bigger questions around the role of academia and whether artificial intelligence is really intelligent.

One possibility is to ask uncomfortable questions that Google and its peers may neglect in their rush to develop ever more powerful algorithms. For example, Berant and his colleagues published an article in 2019 on the biases that creep into natural language systems thanks to the quirks of the individual humans who provide the computer with its initial data. The same is true even with supposedly neutral datasets like Wikipedia. "On the web, if there's a correlation between being a woman and being a nurse," Berant explains, "then of course the models will absorb these biases."

As foundation models are deployed in more and more disciplines — health care, biomedicine, law, education — Berant sees a crucial role for university researchers in tackling issues such as bias, privacy, security and inequity. "There's a lot of interest in academia in these things, which are not about making money for large companies," he says, "but about making sure that these models are deployed in a safe way and that we're aware of both the advantages and the limitations."

That's not to say that the technical challenges of natural language processing have been fully solved, as anyone who has asked Siri, Alexa or Google Translate to tackle anything more than a simple sentence knows. One focus of Berant's current research is the role of reasoning for a question-answering computer program. As presented in a 2021 article by Berant and several colleagues, if you ask a Wikipedia-trained computer whether Aristotle died before the invention of the laptop computer, it won't have any trouble getting the answer, but if you ask it whether Aristotle ever owned a laptop, getting the right answer requires a logical leap. It involves reframing the implicit question as a series of explicit logical steps: When did Aristotle live? When was the laptop invented? Was the former before or after the latter? Training a computer to reliably reason like this remains an ongoing challenge.

A related goal, which Berant tackled in another 2021 paper, is called compositional generalization. Humans are good at putting together pieces of previously learned information to answer questions they've never explicitly seen before. If they know the capitals of every state in the United States, and they know what states border New York, then they can generalize that knowledge to answer the question "What are the capitals of the states that border New York?" Computers, on the other hand, struggle with this. Berant's approach to the problem builds on earlier approaches used by pre-foundation models to break down complex questions into simple components and integrate them into the latest neural network systems.

Where is this all headed? That's a tricky question, Berant acknowledges. The Turing Test was cracked in 2014 when a computer program that pretended to be a thirteen-year-old Ukrainian boy fooled some judges into thinking that it was human. However, no one in the field really believes that makes computers as intelligent, or even as conversationally adept, as humans. "Trying to solve the Turing Test doesn't lead to intelligence, but leads to deceit," Berant says. The same thing happens whenever researchers come up with new benchmarks: computers adopt specific strategies to ace the test, which is done without any need for the general intelligence that the test was designed to elicit.

This problem — what scientists in the field refer to as the "evaluation crisis" — is one of Berant's next targets. "This is something that people in academia, including myself, have been working on a lot." After all, he says, the incredible progress of the last few years toward the ultimate goal of a computer that's fully conversant in natural language raises a crucial question: "How will we know that we got there?" ▲●■

'There's a lot of interest in academia in these things, which are not about making money for large companies, but about making sure that these models are **deployed in a safe way and that we're aware of both the advantages and the limitations.**'