COGNIZANT 20-20 INSIGHTS

Cognizant

Cognitive Computing Systems: How They Learn, How We Make Them Unlearn

When self-learning systems behave in unintended ways, new business roles are needed to identify the source, and then re-train.

INTRODUCTION: KEEPING COGNITIVE SYSTEMS ON-TASK

Real-time detection and response is transforming industries from healthcare to public safety, and from manufacturing to retail. To make urgent decisions faster than humans can, a growing number of organizations across industries are considering cognitive computing systems. According to IDC, worldwide revenues for AI and cognitive systems will grow from nearly $8 billion in 2016 to more than $47 billion in 2020.[1]

Businesses need a plan, however, for when intelligent systems learn in unexpected ways. This is a particular concern given the use cases that IDC expects to grow fastest through 2022, which include public safety, medical diagnosis and quality management. Unexpected actions can have serious consequences.

The fact is, cognitive computing systems do sometimes learn in ways their designers didn't plan. In this way, cognitive computing systems are a lot like humans. A preschooler might learn from cousin Mikey that it's OK to swear. I might learn that kale gives me heartburn when it was really the cheesecake.

Real-world examples of unintended learning abound. Famously, the U.S. Army trained a computer vision system to detect camouflaged enemy tanks. The system performed flawlessly for testers, but not for the Pentagon. The designers eventually solved the mystery: The camouflaged tanks had been photographed on cloudy days and the empty forests on sunny days. The system had learned to associate cloudy days with tanks.[2]

HOW COGNITIVE COMPUTING SYSTEMS LEARN: THREE WAYS

What steps should you take when a cognitive computing system learns something other than what you intended? Like the brain, cognitive computing systems are built from interconnected neural masses. Also like the brain, they're coded only once, developing new behaviors through learning. The maxim is code once, train forever. The remedy is to understand how learning went awry, and then re-train.

This learning happens in one of three ways. For illustration, imagine a system designed to identify possible skin cancer by analyzing images. The simplest type of training is to present hand-chosen input-output pairs. Image A is cancer; image B is not. This is straightforward machine learning. Accuracy remains fairly constant over time but does not improve.

More advanced systems are programmed to periodically consult learning content that developers identify, such as medical image databases. The code might say, "When your error rate increases to X percent, re-learn until your error rate approaches 0 again."

The smartest cognitive computing systems, called self-learning systems, are programmed to seek out new content anytime, anywhere: websites, FTP sites, databases, streaming data from the Internet of Things, PDF documents and so on. Self-learning systems learn continually, leading to better decisions. They need the following capabilities:

- **Advanced computational methods,** such as neural networks, genetic systems, evolutionary systems and Athenic systems.[3]
- **Knowledge of semantics.** These systems can learn the meaning behind language by "reading" documents, for example, or listening to podcasts. If I ask a digital assistant the average kinetic energy of molecules in the room, a self-learning system would take it upon itself to learn that I'm asking for the temperature.
- **Understanding of ontology,** the relationships among things and events based on their properties. An ontology of sports, for example, could include subclasses of "with a ball," "without a ball," "team" and "individual." Other classes, like "Olympic sports" or "timed sports," could be a subclass of the others.

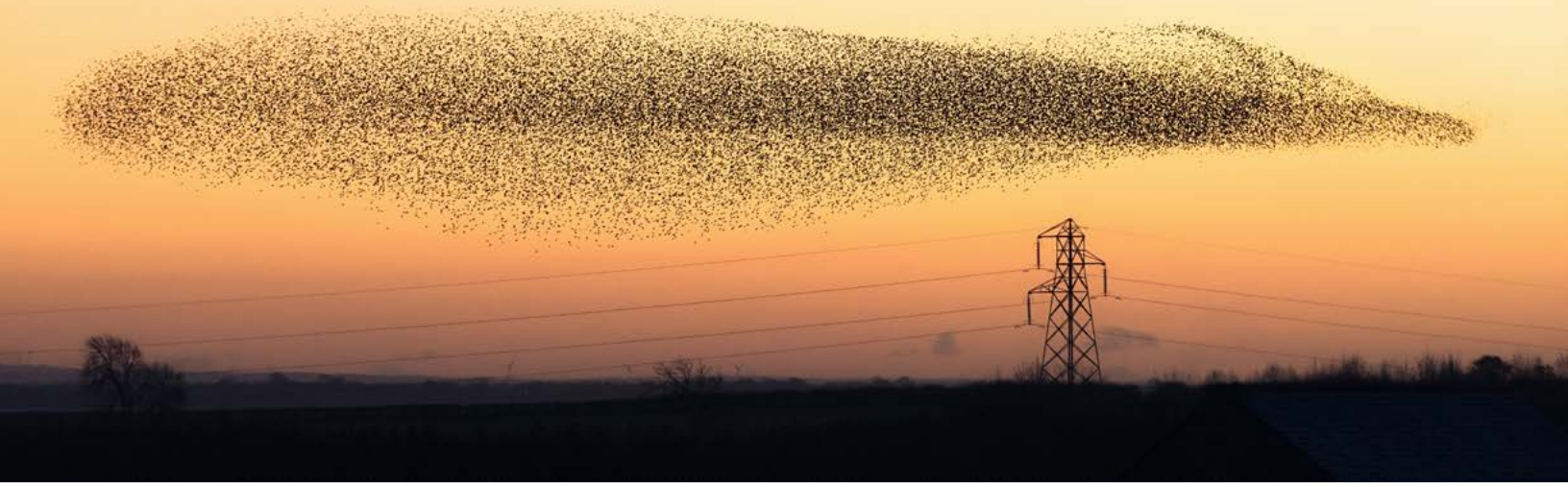
Returning to the cancer example, if you tell a self-learning system, "Find images of skin cancers," it searches every source available over its network connection. Learning from more images increases its diagnostic accuracy. An advanced system might even learn alternate names for the same condition, leading to the discovery of additional images.

UNINTENDED LEARNING HAPPENS

Self-learning systems offer the greatest potential to take over human decisions because they never stop learning. But without a human to curate their sources, these systems are also more likely to learn something that deviates from what their designers intended. Some have developed racial or gender biases, for instance.[4]

Imagine that the self-learning system for identifying skin diseases discovers images in which a tattoo is hiding melanoma. It might erroneously conclude that tattoos are malignancies. False positives would rise, and you wouldn't know why.

The less control you have over a system's learning inputs, the more difficult it is to identify the source of unintended learning. Think about the preschooler. Pinpointing his cousin as the source for his profanity is fairly easy because the parents know most of what the child sees, hears and does. It's far more difficult to identify the source



of unwanted behavior in a teenager, whose influences include multiple circles of friends, as well as books and the Internet.

THE REMEDY: RE-TRAINING

What's the remedy when self-learning systems go off-track? Recoding is not the answer — the maxim in cognitive computing is code once, learn forever. You wouldn't recode a cognitive computing system any more than you'd recode the swearing toddler's DNA.

Updating the database is also not the answer. Cognitive computing systems don't have a database. Neither is adding a rules engine. Static rules can't evolve.

The only antidote to unintended learning is re-training.

Role of the digital psychologist

The first step in re-training a cognitive computing system is finding out what the system learned, and the source. This task requires a new job description: digital psychologist.

The digital psychologist assesses the "patient's" history, noting when the unintended behavior emerged. Activity logs list the sources the system visited within that timeframe. After identifying which source caused the unintended learning, the digital psychologist conducts a controlled learning session. A system that has learned incorrectly that tattoos are cancerous, for example, is shown images of non-malignant tattoos.

Human psychologists also administer tests in their quest to understand the sources of undesired behavior. Unfortunately, tests do not yet exist for cognitive systems. When they do, the digital psychologist might be able to simply ask, "Why do you think this image shows skin cancer?" "Where did you learn that tattoos indicate melanoma?" When available, enterprise-class testing tools will accelerate application development and test and reduce anomalous behaviors.

Role of the digital sociologist

Some cognitive computing systems operate in a decentralized manner: Teams of small, inexpensive, autonomous objects work together to complete a complex task. For example, the U.S. military has successfully tested a swarm of more than 100 autonomous micro-drones.[5] These systems have no centralized controller, and yet the drones can quickly make collective decisions and fly in adaptive formations. This phenomenon — numerous simple entities performing complex behavior as a group — is called emergent behavior. Examples of emergent behavior in nature include migrating geese that take turns flying at the tip of a V formation, and small fish that swim in larger, fish-shaped schools to ward off predators.

Training individual objects in a cognitive computing system to respond to each other in real-time requires a digital sociologist — another new job description. To train tiny bots to remain in proximity to one another, for example, the digital sociologist might develop a series of instructions such as, "If no bots are within one inch, find the nearest pair and move toward them."

As sports fans well know, even if all individuals on a team perform well, the team as a whole may not. The same applies to distributed cognitive computing systems. When undesired emergent behavior appears, the digital sociologist needs to discover the reason and then re-train the individual objects to produce the desired group behavior.

A MORAL EDUCATION

Author Isaac Asimov proposed three laws for robots: Obey humans, but not if it will hurt a human, and not if will hurt the robot. Observing these laws requires complex decision-making. An autonomous vehicle, for example, might need to choose — very quickly — between hitting a pedestrian or crashing into a wall and injuring its passenger. The responsibility for training systems to follow the three ways to act in morally acceptable ways belongs to digital psychologists and digital sociologists.

Like humans, cognitive computing systems can go forth and learn from an ever-growing collection of knowledge. And clearly, business interest is high in adopting these systems to accelerate decisions and improve their accuracy. But cognitive computing systems cannot do this on their own. Intelligent systems will always require humans to observe, guide and retrain them when their behavior goes awry.

FOOTNOTES

[1] "Worldwide Cognitive Systems and AI Revenues Forecast to Surge Past $47 Billion in 2020, According to New IDC Spending Guide," IDC, Oct. 26, 2016, http://www.idc.com/getdoc.jsp?containerId=prUS41878616.

[2] Eliezer Yudkowsky, "Artificial Intelligence as a Positive and Negative Factor in Global Risk," Machine Intelligence Research Institute, 2008, http://intelligence.org/files/AIPosNegFactor.pdf.

[3] "Cognitive Computing: An Antidote for the Data-Intensive, Time-Crunched Age," Cognizant Technology Solutions, April 2017, http://www.rtinsights.com/wp-content/uploads/2017/04/Codex_2732_Cognitive_Computing_RT_Insights_article_1.pdf.

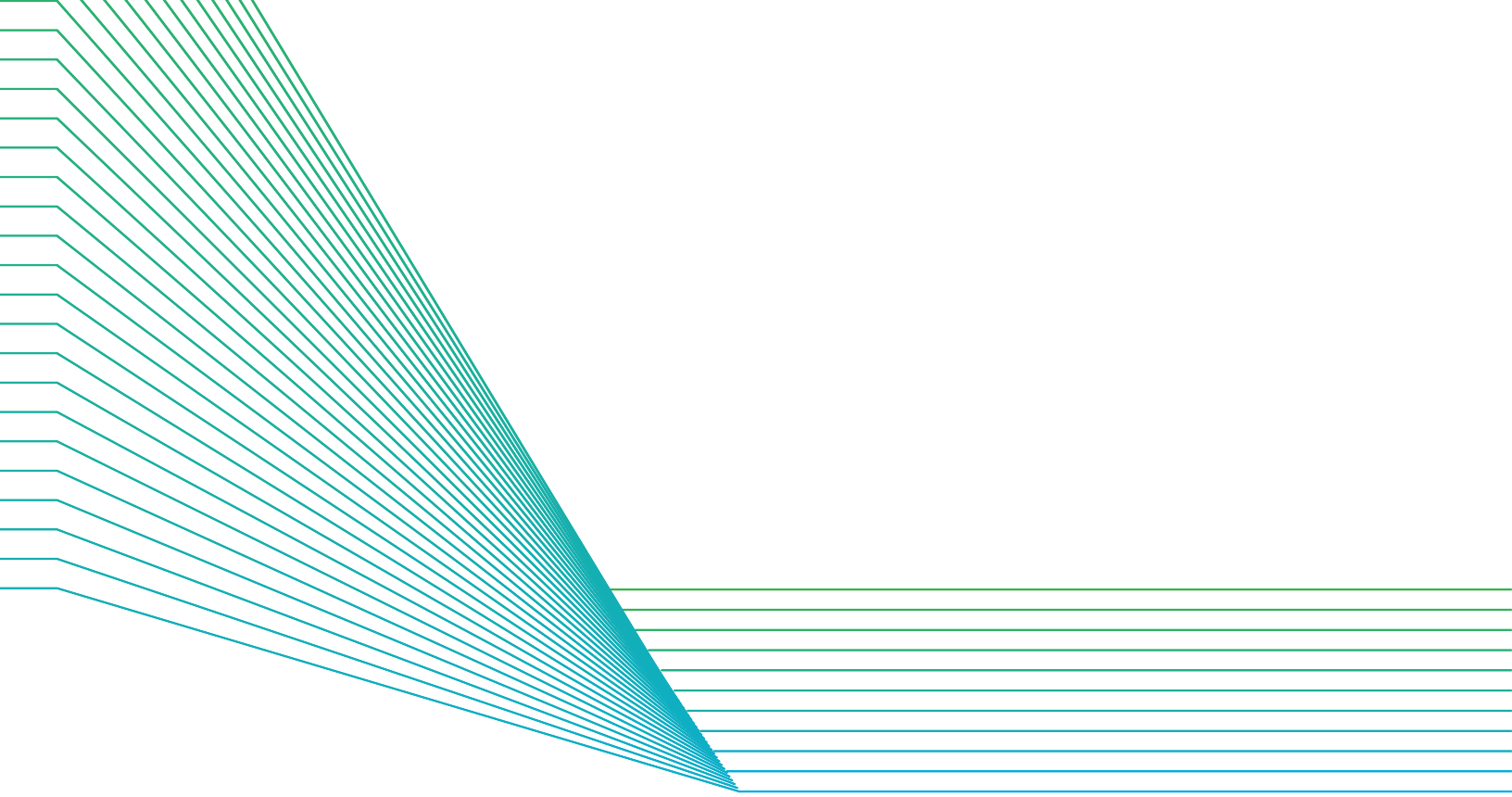
[4] Annalee Newitz, "Princeton Researchers Discover Why AI Became Racist and Sexist," ARS Technica, April 18, 2017, https://arstechnica.com/science/2017/04/princeton-scholars-figure-out-why-your-ai-is-racist/.

[5] "Pentagon Tests Swarm Army of Micro-Drones," Fox 5, Jan. 9, 2017, http://www.fox5ny.com/news/228042592-story#.

ABOUT THE AUTHOR

Jerry A. Smith
Vice-President of Data Sciences, Cognizant

Jerry A. Smith is Vice-President of Data Sciences at Cognizant. He is a practicing data scientist with a passion for realizing business value through enterprise data sciences services. Prior to Cognizant, Jerry was the North American Chief Data Scientist for Capgemini. He has a Ph.D., masters and bachelor of science in computer science, with theoretical and practical experiences in artificial intelligence. Jerry can be reached at Jerry.Smith@cognizant.com | LinkedIn: www.linkedin.com/in/drjerryasmith/.



ABOUT COGNIZANT

Cognizant (NASDAQ-100: CTSH) is one of the world's leading professionalservices companies, transforming clients' business, operating and technology models for the digital era. Our unique industry-based, consultative approach helps clients envision, build and run more innovative and efficient businesses. Headquartered in the U.S., Cognizant is ranked 230 on the Fortune 500 and is consistently listed among the most admired companies in the world. Learn how Cognizant helps clients lead with digital at www.cognizant.com or follow us @Cognizant.

Cognizant

World Headquarters
500 Frank W. Burr Blvd.
Teaneck, NJ 07666 USA
Phone: +1 201 801 0233
Fax: +1 201 801 0243
Toll Free: +1 888 937 3277

European Headquarters
1 Kingdom Street
Paddington Central
London W2 6BD England
Phone: +44 (0) 20 7297 7600
Fax: +44 (0) 20 7121 0102

India Operations Headquarters
#5/535 Old Mahabalipuram Road
Okkiyam Pettai, Thoraipakkam
Chennai, 600 096 India
Phone: +91 (0) 44 4209 6000
Fax: +91 (0) 44 4209 6060

TL Codex 2741