BIG DATA • BIG ANALYTICS • BIG INSIGHTS • (https://www.datanami.com)

(http://tw
/share?te
%E2%80
%99s%2C
%25%20F
url=https
%3A%2F
%2Fwww
ais-1-Home (http://www.datanami.com/)   About (https://www.datanami.com/about/)
problem?
via=datar
Whitepapers (http://www.datanami.com/whitepaper/)   Events (http://www.datanami.com/events/)

f Subscribe (https://www.datanami.com/subscribe/)

(http://w
/sharer.p
%3A%2F
%2Fwww
ais-1-
problem?
t=Exposir
%E2%80
%99s%2C
%25%20F

Sélectionner une langue ▼

Translation Disclaimer (/termsofuse.html#translation)

| Search this site | Search |

Follow Datanami: **f** (http://www.facebook.com/pages
/Datanami/124760547631010)

(http://www.twitter.com/datanami)

(http://www.linkedin.com/groups/Big-Data-News-Network-

4166980) (http://www.datanami.com/feed/)

HOME (HTTP://WWW.DATANAMI.COM/)   FEATURES   SECTORS   APPLICATIONS   TECHNOLOGIES   VENDORS

JOB BANK (HTTP://WWW.DATANAMI.COM/JOB-BANK/)

in
(http://w
/share?u Ad
url=https (https://hpcwire.com/)
%3A%2F
%2Fwww (http://www.enterprisetech.com/)
ais-1-
problem?
title=Exp
%E2%80
%99s%20
%25%20f
(http://www.hpcwire.jp/)

Top News from
(http://w Leading Solution
/submit?l Providers
%3A%2F
%2Fwww (http://www.datanami.com/)
ais-1-
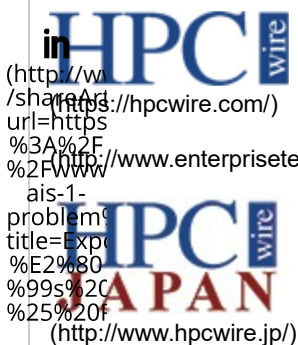
July 25, 2017
Exposing AI's 1% Problem
Alex Woodie

We see the power of artificial intelligence every day: When Netflix recommends a movie you love, when your bank detects fraud in your account, or when Google routes you around a traffic jam. But outside of examples from mammoth companies with millions to spend on data science initiatives, there's a decided lack of AI success among the rest of us.

That's the conclusion that Ali Ghodsi has come to. As the co-founder and CEO of Databricks

(Monte Rego Images/Shutterstock)

| This Just In | Most Read |

July 31, 2017

Attunity Awarded Supplier Status on UK Government's Digital Marketplace (https://www.datanami.com/this-just-in/attunity-awarded-supplier-status-uk-governments-digital-marketplace/)
Strata New York Announces Program Details (https://www.datanami.com/this-just-in/strata-new-york-announces-program-details/)
Carl Data Subsidiary FlowWorks Releases New Application Features (https://www.datanami.com/this-just-in/carl-data-subsidiary-flowworks-

(http://www.databricks.com/) and an adjunct professor at UC Berkeley, Ghodsi has a direct view into the types the AI projects that organizations are embarking on. It turns out, those organizations are struggling mightily, and he wonders why more people aren't talking about it.

"There's a huge gap," Ghodsi tells *Datanami*. "People are not even talking about this problem out there. All the range is about AI and the predictions it can do, but they're not talking about the 99% versus the 1% problem that they have."

Unless your name is Facebook (http://www.facebook.com), Amazon (http://www.amazon.com), Netflix (http://www.netflix.com), or Google (http://www.google.com)– the notorious FANG gang (plus Microsoft (http://www.microsoft.com)) – you're chances of pulling off a successful AI or big data analytics project are slim, according to Ghodsi. "AI has a 1% problem," he says. "There are only about five companies who are truly conducting AI today."

Ghodsi says the AI gap is the result of problems originating in three areas that are critical to big data processing and AI: People, process, and infrastructure.

People Problems

Having the right personnel on hand is critical for success, but it's often overlooked by organizations infatuated with technology and data. Make no mistake: Ghodsi considers data to be king ("It's the data, stupid," he says). But without highly trained data scientist around to turn that raw big data into actionable insight, you're just spinning your wheels, he says.

"Those [FANG] companies have hordes of data scientists, like 10,000 or 20,000 of them. They have PhDs and experts from universities they hired that used to be professors," he says. "But [the rest of the Fortune 2000] say 'We don't have access to Silicon Valley engineers. We just don't have those. There's not enough of them. The people who make huge Silicon Valley engineer salaries over here – that's not the rest of the world.' So how can other companies who don't have the resources to just pour money into hiring 10,000 data scientists – how are they going to do it?"

Companies are banking on software to pick up the slack and help automate data science tasks (https://www.datanami.com/2017/07/21/taking-data-scientist-data-science/) that have largely been executed by people. That's beginning to change – and Databricks, which is the commercial venture behind the Apache Spark software and sells a cloud-based analytical stack that includes Spark — is hoping to ride that wave.

Process Problems

AI programs heavily rely on software programs to find correlations in data, and the majority of that software is open source, which is great. But getting the right pieces of open source software lined up in just the right way is a serious undertaking– and something that Ghodsi says requires the

resources of a FANG (plus Microsoft).

"You can go download 20 pieces of software, hire a bunch of DevOps people whose job is to stitch it all together, and then make sure that the stitches align perfectly so that you don't have any massive scars between the different pieces," he says. "Then as new versions of these softwares come up you better test and make sure they all fit together.

"You can do that," Ghodsi says. "This is essentially what Google and Facebook are doing with their 20,000 engineers….You could also [do that] if you wanted to hire 20,000 data scientist and have them stitch all the open source tools together."

The heavy lifting involved with stitching software together for AI projects is part and parcel of the struggles that companies have encountered (https://www.datanami.com/2017/03/13/hadoop-failed-us-tech-experts-say/) while trying to make use of the Hadoop ecosystem of tools to build data lakes and predictive analytics applications. In many cases, the

Fortune 2000 firms are struggling to close the gap with AI initiatives

differences are semantic, as AI has emerged as the word du jour in reference to advanced analytic use cases.

But Ghodsi does differentiate between big data problems and AI problems, and says organizations need better ways of managing both aspects of the data science challenge.

"Basically, you want to do predictions, and automate those predictions and get insight," he says. "It could be that you're doing traditional machine learning with linear models. It could be that you're doing deep learning using deep neural networks. Whatever you're doing, there are lots of different aspects to this problem that you have to solve, and not all of those are related to machine learning."

Infrastructure Problems

The final element that Ghodsi sees blocking AI's democratic spread to the masses is the infrastructure problem, which includes setting up and managing servers, ensuring the data is secured, and governing access to data scientists or other users.

The FANG companies have a leg up on the rest of the Fortune 2000 because they are cloud companies – managing hardware and granting access to data is what they do best, particularly Google and Microsoft, Ghodsi says.

"Cloud is a very important ingredient here," Ghodsi says. "All the companies mentioned all are cloud companies. They have massive DevOps teams and they've automated a lot of infrastructure. That's where a lot of companies today struggle because they have to go hire DevOps people. And DevOps people don't grow on trees. They're very expensive and they're very hard to hire right now."
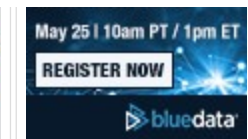
If you thought finding a data scientist was tough, wait until you try to hire a modern DevOps person. But the DevOps people that Ghodsi has in mind are not your grandfather's systems administrators. "They do much more advanced things that sys admins used to do," Ghodsi says. "Those people are very expensive, and hard to find, and a lot of that can be automated in cloud offering. We're huge fans of cloud offerings that automate stuff for you."

The FANG companies (plus Microsoft) use their huge cloud infrastructures to store tons of consumer data that they in turn use to build better consumer products. It's a virtuous cycle, and one that regular enterprises would do well to emulate, Ghodsi says. "You've got to get out of the game of managing all that infrastructure yourself," he says.

It's no coincidence that cloud giants are leaders in big data and AI (Scanrail1/Shutterstock)

Looking Forward

As is the case with most major challenges in this life, there is no silver bullet that will suddenly open the floodgates and put AI within the grasp of the 99% of users who have struggled to make it work.

Ghodsi brings the conversation around to his company's cloud-based offering, dubbed the Unified Analytics Platform, which combines the Apache Spark engine with other elements (such as a data science notebook) to help organizations pursue analytic projects.

"I'm not even saying that Databricks solves that gap," Ghodsi says. "I'm saying we're taking one big leap toward that direction. "

Unless you're one of the FANG companies (plus Microsoft), you're probably not going to have a great AI story to tell. "Why are people not talking about it?" Ghodsi asks rhetorically. "People want to highlight the successes. No one wants to go on the record talking about how difficult it's been to get AI working and we as a company have not have success with it. That's not a sexy story that company leaders are yearning to get out there."

Databricks CEO and co-founder Ali Ghodsi

That's not to say that nobody is succeeding at AI. When pressed, Ghodsi admits that large enterprises are finding success with AI. "That's why they're doubling down," he says. "I just want to highlight that there's a huge gap in the difficulty they're seeing and the challenge they're having, and if you just read TechCrunch every day, you'll see that it's all success story after success story.

Contributors

Alex Woodie

Editor in Chief

George Leopold

Contributing Editor

Steve Conway

IDC

Tiffany Trader

Contributing Editor

"That's my main point," he continues. "Many of these companies have built these data lakes and stored a lot of data in them. But if you ask the companies how successful are you doing predictions on the data lake, you're going to find lots and lots of struggle they're having. We see that all the time. Almost every big company we go to now has built their own data lake that contains a lot of data, but that doesn't mean that they're actually getting value out of those data lakes yet."

Related Items:

Taking the Data Scientist Out of Data Science (https://www.datanami.com/2017/07/21/taking-data-scientist-data-science/)

Hadoop Has Failed Us, Tech Experts Say (https://www.datanami.com/2017/03/13/hadoop-failed-us-tech-experts-say/)

Anatomy of a Hadoop Project Failure (https://www.datanami.com/2017/03/17/anatomy-hadoop-project-failure/)

Share this:

Tweet   **Share**   **f Share**   G+

reddit this!   (//www.reddit.com/submit?url=https://www.datanami.com/2017/07/25/exposing-ais-1-problem/)

Tags: AI (https://www.datanami.com/tag/ai/), Ali Ghodsi (https://www.datanami.com/tag/ali-ghodsi/), artificial intelligence (https://www.datanami.com/tag/artificial-intelligence/), big data (https://www.datanami.com/tag/big-data/), databricks (https://www.datanami.com/tag/databricks/), deep learning (https://www.datanami.com/tag/deep-learning/), FANG (https://www.datanami.com/tag/fang/), machine learning (https://www.datanami.com/tag/machine-learning/)

Featured Events

Artificial Intelligence Conference San Francisco 2017 (https://www.datanami.com/event/artificial-intelligence-conference-san-francisco-2017-2/)

September 17 - September 20

San Francisco CA United States

Strata Data Conference New York 2017 (https://www.datanami.com/event/strata-data-conference-new-york-2017/)

September 25 - September 28

New York United States

Chief Data & Analytics Officer Week 2017 (https://www.datanami.com/event/chief-data-analytics-officer-week-2017/)

September 25 - September 28

DCD >Colo + Cloud (https://www.datanami.com/event/dcd-colo-cloud/)

September 26

Dallas TX United States

Social share sidebar

Twitter (http://tw /share?te %E2%80 %99s%20 %25%20F url=https %3A%2F %2Fwww ais-1- problem via=datar

Facebook (http://w /sharer.p %3A%2F %2Fwww ais-1- problem t=Exposir %E2%80 %99s%20 %25%20F

LinkedIn (http://w /shareArt url=https %3A%2F %2Fwww ais-1- problem title=Exp %E2%80 %99s%20 %25%20F

Reddit (http://w /submit?u %3A%2F %2Fwww ais-1-