CEOCFO Magazine

DeepCode is helping Developers Write Better Code providing a Platform with Machine Learning Algorithms and AI that Identifies Bugs Before they Happen

Boris Paskalev
CEO & Co-Founder

DeepCode

Contact:
CEOCFO Magazine
570-851-1745

Interview conducted by:
Lynn Fosse, Senior Editor
CEOCFO Magazine

CEOCFO: *Mr. Paskalev, what is the concept behind DeepCode?*
Mr. Paskalev: The concept behind DeepCode is pretty simple. In short, we are automatically crowdsourcing all the knowledge of the global development community. Understanding software problems that were fixed and exactly how they were fixed allows us to build this unprecedented knowledge base that prevents every DeepCode user from repeating all known mistakes. DeepCode identifies bugs before they happen to fulfill developers' true coding potential and deliver quality products, faster.

CEOCFO: *How do you get all of the information in one place?*
Mr. Paskalev: We use a proprietary analysis, which is obviously the hard piece, but the main source of data is coming from open source repositories. There are billions of lines of code and hundreds of millions of fixes that already exist out there. Therefore, we actually look at the history of every single fix ever made since the very first line of software code was written. We collect that knowledge and continuously update it as the development community and different languages evolve.

CEOCFO: *When you are looking at all of the information or when the programmer is looking, how do you know when something was an old answer, but now there is a better answer? What has changed historically?*
Mr. Paskalev: Pretty much as we said, we track the versions from the very beginning until very recently. If you imagine the knowledge that we collect as a sphere, on the outside of the sphere we have the latest and greatest way to fix a particular problem, and as you go deeper you see the history of the developers' knowledge. Therefore, we not only see what the previous versions are but exactly when the development community changed their mind on how to fix a specific problem.

CEOCFO: *DeepCode offers a proprietary technology. How are you different from other solutions?*
Mr. Paskalev: We have the most advanced program analysis. We actually treat the software code as a white box. We represent every single object and every single interaction between the object and build a map/representation of each program.

We also analyze the software on a semantic level and the interactions of different objects as they propagate through producers, functions, APIs, etc. For training/learning, this analysis happens over billions of lines of code and millions of commits. We also run an ever-growing range of predictive algorithms that infer the specification and functional use of various functions and pieces of code. This combination of technologies is really the key differentiator of DeepCode's software platform. Software is pretty complex as a structure—there are no obvious standards/patterns that you can find just by reading the code. This is why we have to convert code into a machine learnable format, where we can use

powerful machine learning algorithms to look at the multitude of data and then extract knowledge for rules, patterns and up-to-date solutions to coding issues.

CEOCFO: *Is DeepCode available today? Are you still in development?*
Mr. Paskalev: It is fully available at www.deepcode.ai. We are actually free for the open source community, so everyone can start using DeepCode in less than a minute. They just have to log in and all of their repositories will automatically be subscribed and quickly analyzed. We are also free for educational purposes and small enterprises of up to thirty developers.

CEOCFO: *What is your business model? How will you be making money?*
Mr. Paskalev: Number one, if you have a large enterprise, we assume that you are profitable from those software developers, and this is when we charge for the service, especially if you have an on-premise solution. On-premise almost always requires some manual setup and then targeted support. In the long-term, the main revenue will come from additional services that we are building on top of DeepCode's software platform. We have an impressive list of new services planned that will help developers and automate as much of their manual work as possible.

CEOCFO: *How does someone use DeepCode?*
Mr. Paskalev: It is very simple, depending on where you store your code, but usually, that is in a Git (eighty percent of the code in the world is in Gits today). You just log in with your user name and password from your GitHub or Bitbucket account and then immediately you see all of your repositories and subscribe them for continuous monitoring or run a one-time analysis for the issues identifiable today.

> **"DeepCode identifies bugs before they happen to fulfill developers' true coding potential and deliver quality products, faster."- Boris Paskalev**

CEOCFO: *How does someone have to ask for the information?*
Mr. Paskalev: That is a very good question about knowledge. The best part is that developers do not require any setup or knowledge to start using the system. We automatically analyze code and highlight the pieces of the code with potential problems. We not only do that, but we actually explain to developers what the problem is and show examples of how other people have fixed the very same or similar problem in a totally different context. Therefore, even if a developer does not understand the problem, they can still see a solution so they can quickly fix it.

CEOCFO: *How do you keep up to date? I assume you are constantly monitoring data?*
Mr. Paskalev: Yes, we never stop learning. We actually monitor every single new line of code being written. Our back-end platform continuously ingests new changes, analyzes them and assimilates them into the DeepCode knowledge base, which is continuously used to serve our users.

CEOCFO: *How are you spreading the word about DeepCode?*
Mr. Paskalev: We have not so far. This was the first week that we actually published any official articles as part of our fundraising, so we have not done much. We are actually experimenting with various channels to share the message of DeepCode with the global development community. It has mainly just been word of mouth and people hearing about DeepCode and asking questions.

CEOCFO: *How will you be using the money from your recent funding?*
Mr. Paskalev: The main focus for us is to further extend the platform. We will expand the team to support new languages and integrations, clearly develop the technology further so it gets better and better as we go, and allow the creation of new services, as I mentioned. Those are the extra things that developers are looking for and are what will help them a lot in their development workflow.

CEOCFO: *What enhancements are in the works?*
Mr. Paskalev: The measurement of quality that we track is a bit of machine learning lingo, but it is called recall, which means how many actual issues we can find. We are way ahead of any existing system in terms of what we can find and how accurately we can find them. This is really our main focus.

There is no great definition of what all the problems in software are, but our goal is to continuously increase what we can find. Our ultimate goal is to save developers' time so they do not have to use other developers or their own time to manually review code for known issues. These issues should be found automatically.

CEOCFO: *What were some of the challenges in putting this all together?*
Mr. Paskalev: The main challenge, as I alluded to in the beginning, is to be able to create a representation that really understands code and simulates the program with extremely high precision. Normally you would need real developers looking at the code for quite some time to understand and grasp what it is all about. This is the area that we continue to develop and maintain a representation that is machine learnable at scale.

It has to be fast and efficient in addition to accurate. This is the main backbone for the DeepCode software platform. Then the rest is all about how we actually deliver/package that to the users.

CEOCFO: *Have similar approaches been tried?*
Mr. Paskalev: There are many approaches out there! If I am to guess and go by what I have looked at, there are more than one hundred different tools that are trying similar types of automation. However, they are using totally different and often simplistic methods or are highly focused on only one area/language.

But in all systems with any significance, there is a need to manually create rules for each language and for each variance of a problem. We recently looked at date-time formatting, due to a customer request. To catch all variations of misrepresenting that you need close to 100 manually written rules, multiplied by the number of languages supported. You might be looking at weeks of work, and even then, it's unlikely that you'll cover all of the possible issues. The difference is that no one has really gone deep into being able to learn continuously and automatically from billions of lines of code.

CEOCFO: *When did you and your cofounders get the idea? How did things start?*
Mr. Paskalev: The reason that we started DeepCode is the work of Dr. Veselin Raychev and Prof. Dr. Martin Vechev (co-founders of DeepCode) at ETH Zurich University, which is often referred to as the MIT of Europe. They actually spent about six years in research there on the topics of learning from big code and program analysis. They also launched a number of open source services that are pretty widely used today. Our Co-Founder and CTO, Dr. Raychev, published his PhD and was one of the best-awarded PhDs in the world at that time. It became obvious that there was a huge gap in the market and the research could revolutionize software development as we know it. That was where the idea came for starting DeepCode as a company.

CEOCFO: *What surprised you so far through the process?*
Mr. Paskalev: What has surprised me is the complexity of the platform, the changes in research, and how many different modules have to be built in order to enable this. That surprised me two years ago when I started looking into it in detail, and it still surprises me today how many new things we have to invent every week to further the platform. Our team remains super excited and insists on publishing some of the interesting findings. At the current rate, the cutting-edge topics incorporated into the platform will likely double or triple over the next one to two years.

CEOCFO: *Why look at DeepCode right now?*
Mr. Paskalev: People should pay attention to DeepCode because it is changing the industry. It is almost like the impact that natural language processing, automatic translations, and automated spell check have had.

There is no doubt that it will become a standard because software development is getting more and more complex, developers are hard to find, and there are many more issues coming up. The programming languages are very dynamic and growing in number, so it is inevitable that an autonomous, language-independent platform will be the foundation of most of the developer tools we use. As of right now, no one else is really working on this technology, so that makes us uniquely positioned in this space, and our AI Code Review service is already proving that over existing tools.