

# IBM veröffentlicht Testdaten für KI-gestützte Softwaremodernisierung

Originalartikel

Backup

<html> <header class=„article-header“><h1 class=„articleheading“>IBM veröffentlicht Testdaten für KI-gestützte Softwaremodernisierung</h1><div class=„publish-info“> Robert Lippert</div></header><figure class=„aufmacherbild“><img src=„[https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de\\_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg](https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg)“ srcset=„[https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de\\_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg\\_700w](https://heise.cloudimg.io/width/700/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg_700w), [https://heise.cloudimg.io/width/1050/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de\\_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg\\_1050w](https://heise.cloudimg.io/width/1050/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg_1050w), [https://heise.cloudimg.io/width/1500/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de\\_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg\\_1500w](https://heise.cloudimg.io/width/1500/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg_1500w), [https://heise.cloudimg.io/width/2300/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de\\_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg\\_2300w](https://heise.cloudimg.io/width/2300/q75.png-lossy-75.webp-lossy-75.foil1/www-heise-de_imgs/18/3/1/0/5/4/1/4/KI-02-9b952b62e091f175.jpeg_2300w)“ alt=„IBM veröffentlicht Testdaten für KI-gestützte Modernisierung“ class=„img-responsive“ referrerpolicy=„no-referrer“><figcaption class=„akwa-caption“>(Bild:&#160;<a href=„<https://pixabay.com/de/users/geralt-9301/>“ target=„\_blank“ rel=„external noopener“>Gerd Altmann</a>, gemeinfrei)</figcaption></figure><p><strong>Umfangreich und divers: IBMs Project CodeNet liefert Testdaten für die Entwicklung künstlicher Intelligenzen für die Softwaremodernisierung.</strong></p><p>Rund 14 Millionen Codebeispiele und über 500 Millionen Zeilen Quellcode zu 55 unterschiedlichen Programmiersprachen &#8211; mit Project CodeNet hat IBM jetzt ein umfangreiches Set an Testdaten für die Entwicklung künstlicher Intelligenzen für die Softwareentwicklung und -pflege veröffentlicht. Es soll die Grundlage schaffen für automatisierte Code-Korrekturen und Code-Übersetzungen und damit die KI-gestützte Modernisierung komplexer, monolithischer Anwendungen erleichtern. Insbesondere sind in dem Set Daten für populäre Sprachen wie C, C++, Java und Python enthalten.</p><p>Als Demonstration für die Leistungsfähigkeit des Datensets berufen sich die Forscher bei IBM auf ein nicht nachgewiesenes Referenzprojekt, in dem sie einen auf ein Jahr ausgelegten Migrationsprozess mithilfe des KI-Stacks aus CodeNet in nur vier Wochen umsetzen konnten. Dabei wurde eine Legacy-Applikation in über 25 Cloud-native Microservices zerlegt.</p><h3 class=„subheading“ id=„nav\_programmiersprac0“>Programmiersprachen schwer zu übersetzen</h3><p><a href=„<https://www.research.ibm.com/blog/codenet-ai-for-code>“ rel=„external noopener“ target=„\_blank“><strong>In einem Blogbeitrag [1]</strong></a> gehen die Wissenschaftler nach auf die Herausforderung ein, dass Programmiersprachen sich nicht ohne Weiteres regelbasiert übersetzen ließen. Gerade in großen Codebasen erschließen sich der Kontext einzelner Statements erst über mehrere Bibliotheken hinweg. Mit bestehender Technologie ließen sich dabei nur bis zu 60 Prozent eines Programms übersetzen; mit ihrem KI-Stack aus CodeNet versprechen die Forscher an dieser Stelle eine deutliche Verbesserung.</p><p>Neben reinen Codebeispielen liefert IBM mit CodeNet auch diverse Metadaten und Annotationen, von Codegrößen und Speicherverbrauch bis hin zu CPU-Laufzeiten, was darauf aufbauenden KIs eine Unterscheidung zwischen Akzeptanz- und Fehlerkriterien erleichtern soll.</p><h3 class=„subheading“ id=„nav\_das\_imagenet1“>Das ImageNet der Softwaremodernisierung?</h3><p>Nach Einschätzung von IBM kann CodeNet sich

langfristig für die Entwicklung künstlicher Intelligenzen ähnlich erfolgreich erweisen, wie ImageNet es im Bereich des maschinellen Sehens vorgemacht hat. Die Praxis wird zeigen, ob CodeNet dabei auch eine vergleichbare Fehlerquote aufweist [1]; erst jetzt hat eine Studie zweier Wissenschaftler des MIT zusammen mit einem Amazon-Mitarbeiter [die Zuverlässigkeit von Machine-Learning-Datensätzen untersucht \[2\]](https://www.heise.de/news/KI-Studie-findet-viele-falsche-Beschreibungen-in-Machine-Learning-Datensaetzen-6000541.html). In dem für die Bildklassifikation häufig verwendeten ImageNet-Datensatz hat die Prfung eine Fehlerquote von 5,83 Prozent ausgemacht. Das CodeNet-Datenset [steht auf GitHub \[3\]](https://github.com/IBM/Project_CodeNet) unter der freien Apache-Lizenz zur Verfügung. Begleitend finden sich hier auch Tools unter anderem für die statistische Auswertung oder die Versetzung zwischen beliebten Datenformaten. ()

URL dieses Artikels: [Links in diesem Artikel:](https://www.heise.de/-6045282)

[1] <https://www.research.ibm.com/blog/codenet-ai-for-code>

[2] <https://www.heise.de/news/KI-Studie-findet-viele-falsche-Beschreibungen-in-Machine-Learning-Datensaetzen-6000541.html>

[3] [https://github.com/IBM/Project\\_CodeNet](https://github.com/IBM/Project_CodeNet)

</p><p class="printversioncopyright"><em>Copyright &#169; 2021 Heise Medien</em></p>

From:  
<https://schnipsl.qgelm.de/> - Qgelm

Permanent link:  
<https://schnipsl.qgelm.de/doku.php?id=wallabag:wb2ibm-verffentlicht-testdaten-fr-ki-gesttzte-softwaremodernisierung>

Last update: 2025/06/27 11:17

