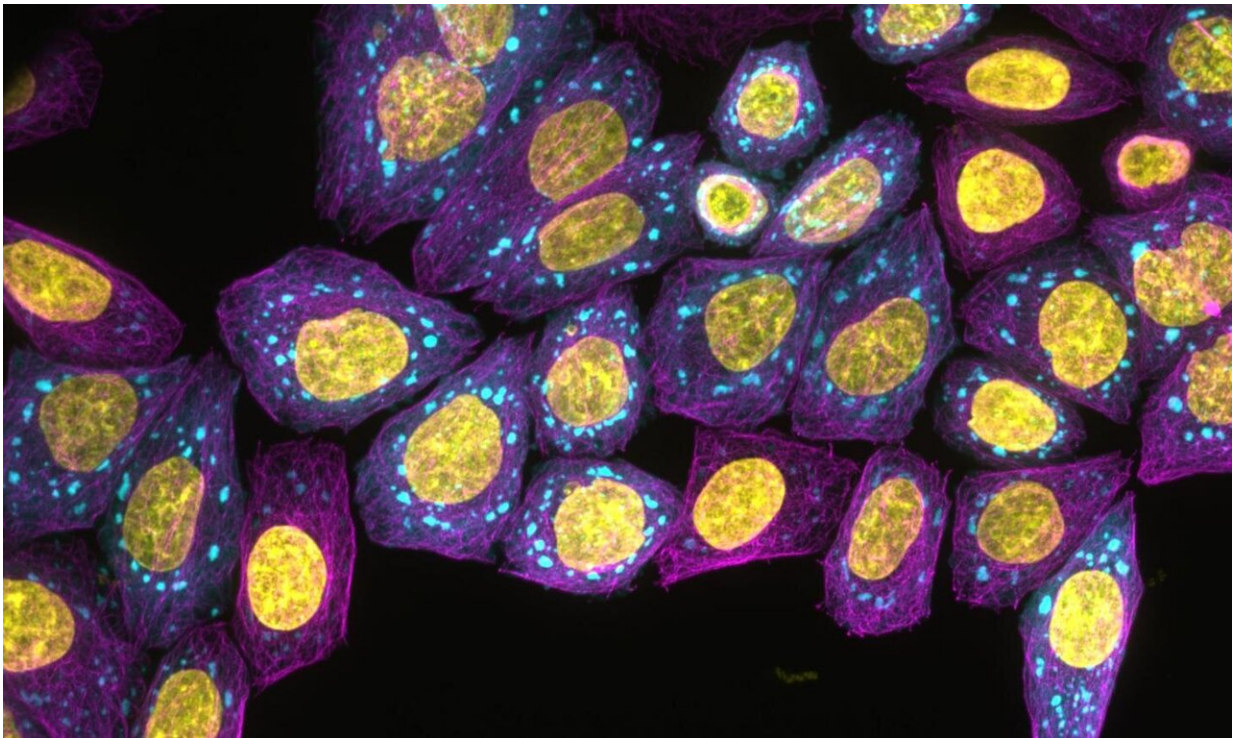


# Artificial Intelligence could 'crack the language of cancer and Alzheimer's'

April 8 2021

---



Fluorescence microscopy image of protein condensates forming inside living cells. Credit: Weitz lab, Harvard University

Powerful algorithms used by Netflix, Amazon and Facebook can 'predict' the biological language of cancer and neurodegenerative diseases like Alzheimer's, scientists have found.

Big data produced during decades of research was fed into a computer [language](#) model to see if artificial intelligence can make more advanced discoveries than humans.

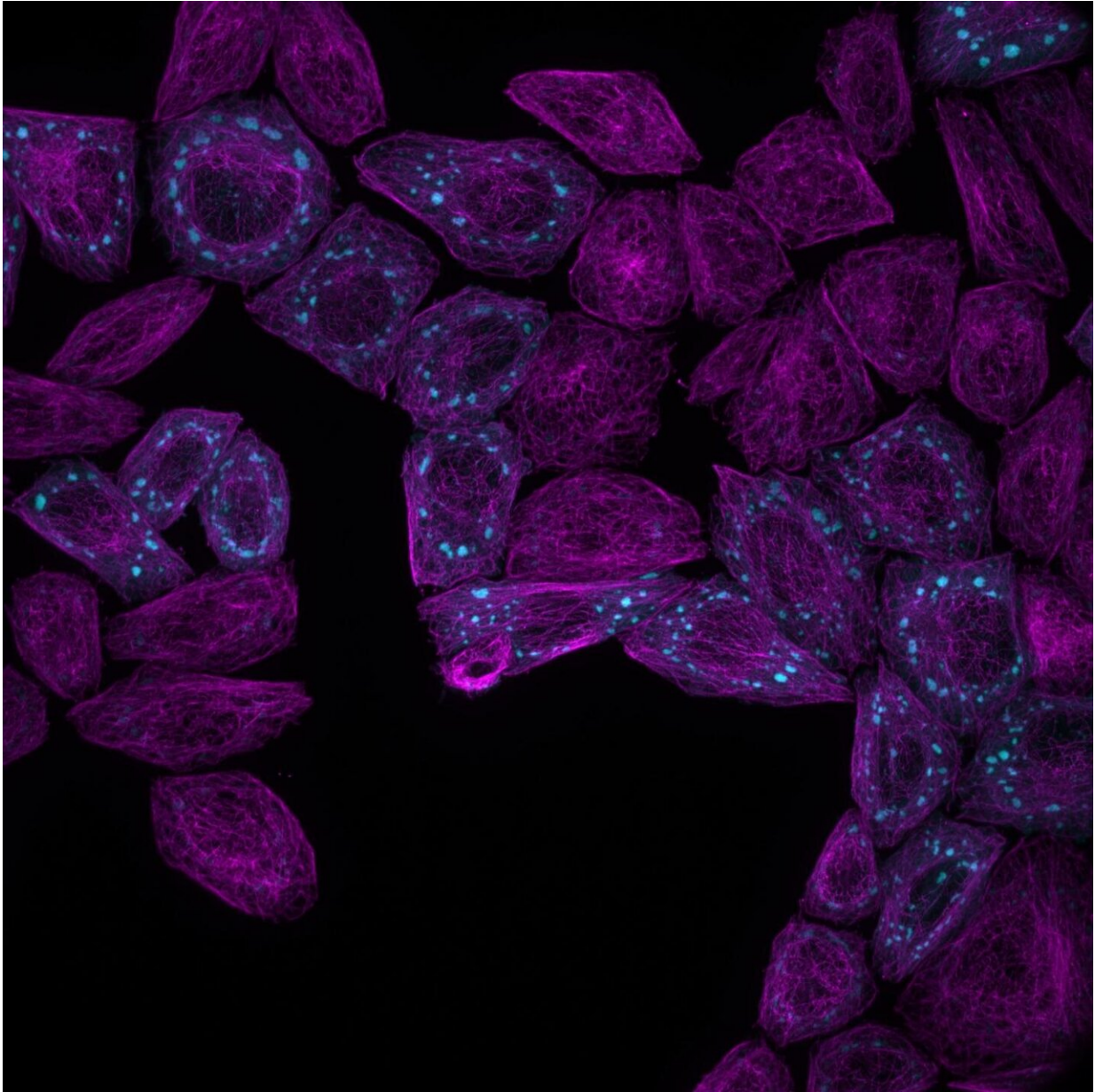
Academics based at St John's College, University of Cambridge, found the machine-learning technology could decipher the 'biological language' of cancer, Alzheimer's, and other [neurodegenerative diseases](#).

Their ground-breaking study has been published in the scientific journal *PNAS* today and could be used in the future to 'correct the grammatical mistakes inside cells that cause disease'.

Professor Tuomas Knowles, lead author of the paper and a Fellow at St John's College, said: "Bringing machine-learning technology into research into neurodegenerative diseases and cancer is an absolute game-changer. Ultimately, the aim will be to use [artificial intelligence](#) to develop targeted drugs to dramatically ease symptoms or to prevent dementia happening at all."

Every time Netflix recommends a series to watch or Facebook suggests someone to befriend, the platforms are using powerful machine-learning algorithms to make highly educated guesses about what people will do next. Voice assistants like Alexa and Siri can even recognize individual people and instantly 'talk' back to you.

Dr. Kadi Liis Saar, first author of the paper and a Research Fellow at St John's College, used similar machine-learning technology to train a large-scale language model to look at what happens when something goes wrong with proteins inside the body to cause disease.



Fluorescence microscopy image of protein condensates forming inside living cells. Credit: Weitz lab, Harvard University

She said: "The human body is home to thousands and thousands of proteins and scientists don't yet know the function of many of them. We asked a [neural network](#) based language model to learn the language of

proteins.

"We specifically asked the program to learn the language of shapeshifting biomolecular condensates—droplets of proteins found in cells—that scientists really need to understand to crack the language of biological function and malfunction that cause cancer and neurodegenerative diseases like Alzheimer's. We found it could learn, without being explicitly told, what scientists have already discovered about the language of proteins over decades of research."

Proteins are large, complex molecules that play many critical roles in the body. They do most of the work in cells and are required for the structure, function and regulation of the body's tissues and organs—antibodies, for example, are a protein that function to protect the body.

Alzheimer's, Parkinson's and Huntington's diseases are three of the most common neurodegenerative diseases, but scientists believe there are several hundred.

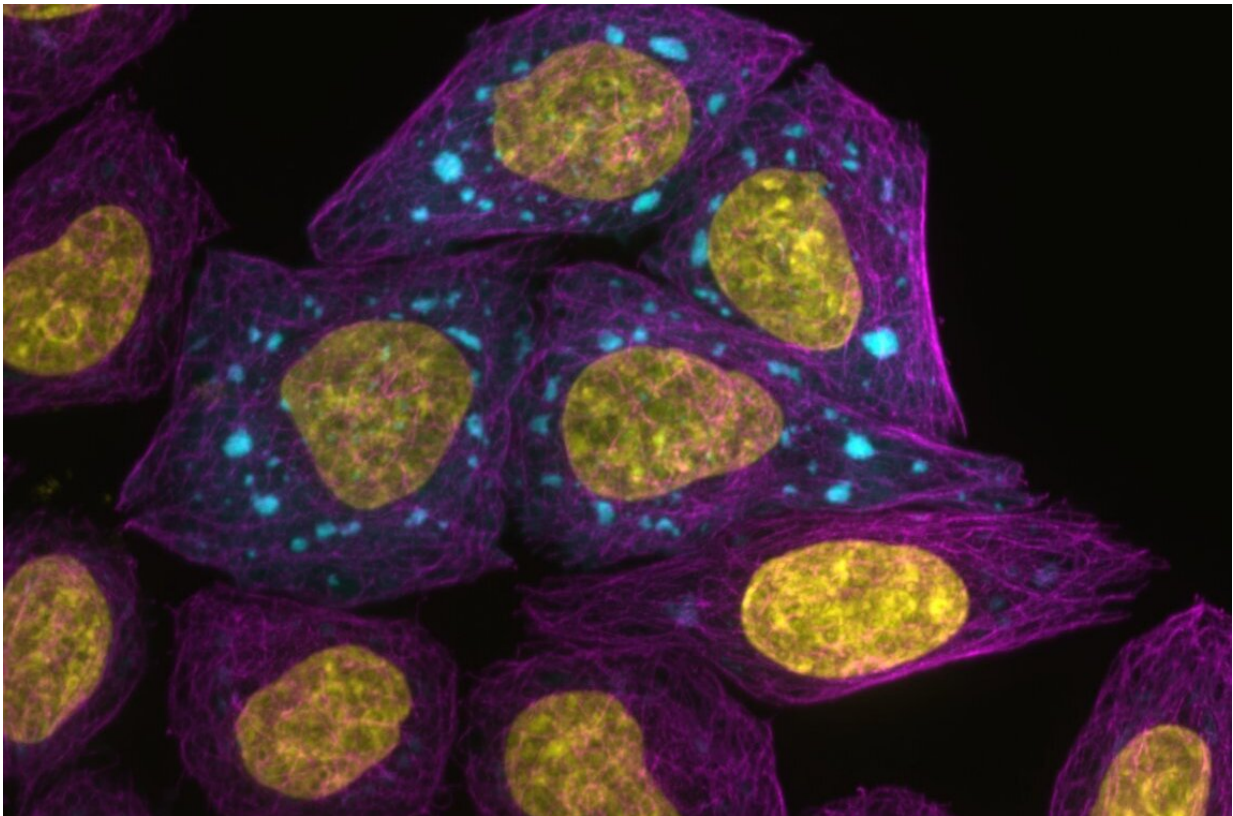
In Alzheimer's disease, which affects 50 million people worldwide, proteins go rogue, form clumps and kill healthy nerve cells. A healthy brain has a quality control system that effectively disposes of these potentially dangerous masses of proteins, known as aggregates.

Scientists now think that some disordered proteins also form liquid-like droplets of proteins called condensates that don't have a membrane and merge freely with each other. Unlike protein aggregates which are irreversible, protein condensates can form and reform and are often compared to blobs of shapeshifting wax in lava lamps.

Professor Knowles said: "Protein condensates have recently attracted a lot of attention in the scientific world because they control key events in



the cell such as gene expression—how our DNA is converted into proteins—and protein synthesis—how the cells make proteins.



Fluorescence microscopy image of protein condensates forming inside living cells. Credit: Weitz lab, Harvard University

"Any defects connected with these protein droplets can lead to diseases such as cancer. This is why bringing natural language processing technology into research into the molecular origins of [protein](#) malfunction is vital if we want to be able to correct the grammatical mistakes inside cells that cause disease."

Dr. Saar said: "We fed the algorithm all of data held on the known

proteins so it could learn and predict the language of proteins in the same way these models learn about human language and how WhatsApp knows how to suggest words for you to use.

"Then we were able ask it about the specific grammar that leads only some proteins to form condensates inside cells. It is a very challenging problem and unlocking it will help us learn the rules of the language of disease."

The [machine-learning technology](#) is developing at a rapid pace due to the growing availability of data, increased computing power, and technical advances which have created more powerful algorithms.

Further use of machine-learning could transform future cancer and neurodegenerative [disease](#) research. Discoveries could be made beyond what scientists currently already know and speculate about diseases and potentially even beyond what the human brain can understand without the help of machine-learning.

Dr. Saar explained: "Machine-learning can be free of the limitations of what researchers think are the targets for scientific exploration and it will mean new connections will be found that we have not even conceived of yet. It is really very exciting indeed."

The network developed has now been made freely available to researchers around the world to enable advances to be worked on by more scientists.

**More information:** Kadi L. Saar et al, Learning the molecular grammar of protein condensates from sequence determinants and embeddings, *Proceedings of the National Academy of Sciences* (2021). [DOI: 10.1073/pnas.2019053118](https://doi.org/10.1073/pnas.2019053118)

Provided by University of Cambridge

Citation: Artificial Intelligence could 'crack the language of cancer and Alzheimer's' (2021, April 8) retrieved 5 May 2024 from <https://medicalxpress.com/news/2021-04-artificial-intelligence-language-cancer-alzheimer.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.