PRIZE ESSAY

GRAND PRIZE WINNER

Michael A. Skinnider

Michael Skinnider received his undergraduate degree from McMaster University and a MD/PhD from the University of British Columbia, during which time he was also a visiting PhD student at the École Polytechnique Fédérale de Lausanne. He then started his laboratory in the Lewis-Sigler Institute for Integrative Genomics and the Ludwig Institute for Cancer Research at Princeton University. His research focuses on the application of machine learning to problems in biology, chemistry, and medicine. **www.science.org/doi/10.1126/science.adk8626**

NOMIS & Science YOUNG EXPLORER AWARD

New designer drugs present unknown risks to users and treatment challenges to first responders.

SOCIAL SCIENCES

Hallucinating hallucinogens

Fighting the designer drug epidemic with generative AI

By **Michael A. Skinnider**[1,2]

Early on the morning of 12 July 2016, paramedics were called to the scene of a mass casualty event in Brooklyn (*1*). Bystanders had reported that multiple individuals were displaying unusual behavior that they described as "zombie-like." Investigators quickly established that all the affected individuals had smoked a packet of "herbal products" marketed as "AK-47 24 Karat Gold." But this was not a typical drug overdose: Most recreational drugs, after all, do not cause users to stagger mechanically through the street, moaning and staring blankly. What "herbal product" could cause such alarming symptoms?

Forensic scientists suspected the packet contained a completely new "designer drug" that had just emerged on the illicit market. Over the following weeks, a team of forensic chemists worked feverishly to establish the identity of the drug: a potent synthetic cannabinoid named AMB-FUBINACA. This compound had originally been developed by Pfizer but was later abandoned and never tested in humans. In vitro data suggested that AMB-FUBINACA was 85 times more potent than Δ^9-tetrahydrocannabinol, the principal psychoactive constituent of cannabis—a level of potency that could account for its zombie-like effects (*2*).

AMB-FUBINACA is just one of a wave of designer drugs that have reshaped the illicit drug market over the past two decades. Conventional drugs of abuse such as cocaine or methamphetamine still dominate the market, but enterprising chemists have realized that with slight chemical modifications to these drugs, they can create new derivatives that are completely legal (*3*).

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [2]Ludwig Institute for Cancer Research, Princeton University, Princeton, NJ, USA. Email: skinnider@princeton.edu

But because most of these derivatives have never been tested in humans, they can have unpredictable and even fatal effects. This situation presents enormous challenges to first responders, who must treat completely new toxidromes, as well as public health officials, who must determine how to respond to outbreaks caused by uncharacterized molecules (*4*).

If these drugs could be identified faster, emergency physicians could provide better treatment to intoxicated patients, and public health officials could better guide responses to outbreaks. But identifying a new designer drug is no small feat. The most commonly used analytical method—mass spectrometry—requires a pure sample of the drug of interest for comparison. Yet pure samples are rarely available for drugs that have just emerged on the illicit market, and without pure samples, the identification of a new drug can take months.

As a MD/PhD student, I saw firsthand how patients could present with devastating symptoms of designer drug intoxications, but emergency physicians had few options to treat them. I wondered whether artificial intelligence (AI) could help. Specifically, I asked whether AI could automatically elucidate the chemical structures of new designer drugs from mass spectrometry data. Scientifically, this was a tall order; structure elucidation of unknown molecules with mass spectrometry is widely seen as an impossible task. But advances in generative AI made me optimistic. A new class of machine-learning models, called language models, had just begun to attract chemists' attention (*5*). Originally designed to parse human language, these models could instead be trained on textual representations of chemical structures called SMILES (simplified molecular input line entry system) (*6*). Once trained, language models could generate entirely new chemical structures that resembled the molecules in the training set. This phenomenon has been referred to as "hallucination" because the generated SMILES do not correspond to known chemical structures but instead reflect what the neural network views as plausible extensions of the training set (*7*). I thought that these models might overcome one of the primary challenges in structure elucidation: namely, generating unknown molecules to match the mass spectrometry data.

To test this hypothesis, I first had to overcome a major technical challenge. Previous language models had been trained on millions of chemical structures. But a crowdsourced effort to assemble the structures of all known designer drugs had come up with just 1753 molecules (*8*). It was unknown whether language models could even learn from so little data.

Unexpectedly, my colleagues and I found that with a slight tweak, language models could learn from as few as 1000 molecules (*9*). The key insight was to change how the molecules were represented—specifically, that the same chemical structure can be represented with many different SMILES (*10*). We found that this redundancy could provide a powerful mechanism for data augmentation: By enumerating many different SMILES for each molecule, the model could be tricked into thinking it was learning from a larger dataset of molecules.

Using this data augmentation trick, we successfully trained a language model on the structures of all 1753 known designer drugs and showed that this model could generate entirely new molecules whose structural properties were indistinguishable from those of known drugs (*11*). Intriguingly, these molecules were not all equally likely to be generated by the model. In a sample of 1 billion generated molecules, most chemical structures appeared just once or twice, but a tiny minority were generated tens of thousands of times.

This observation raised an exciting possibility: Could the model be learning to statistically anticipate the designer drugs that were most likely to emerge on the illicit market next? With the help of the Danish national forensic laboratory, we tested this hypothesis on a designer drug that had just been discovered: a new derivative of the street drug PCP (phencyclidine) named deoxymethoxetamine. I showed that given only the mass of this drug—one of the cheapest and easiest properties to measure experimentally—the language model could correctly elucidate deoxymethoxetamine's entire chemical structure (*11*).

As a final step, we integrated the language model's predictions with mass spectrometric data. We showed that combining these two sources of information allowed the model to elucidate the chemical structures of 40 new designer drugs (*11*). These results suggest that language models could dramatically accelerate the pace at which emerging drugs are identified.

I have now applied this technology to tens of thousands of patient samples and used it to discover several new designer drugs, such as a new analog of fentanyl that emerged last year. Currently, I am working with the British Columbia Centre for Disease Control to implement this AI technology in routine clinical practice to automatically discover new drugs as soon as they are introduced into the population. Ultimately, my dream is that first responders, emergency physicians, and public health officials will all be able to take advantage of generative AI to make more informed decisions when treating patients and managing outbreaks. ■

REFERENCES AND NOTES

1. A. J. Adams *et al.*, *N. Engl. J. Med.* **376**, 235 (2017).
2. S. D. Banister *et al.*, *ACS Chem. Neurosci.* **7**, 1241 (2016).
3. E. Underwood, *Science* **347**, 469 (2015).
4. E. Gebel Berg, *ACS Cent. Sci.* **2**, 363 (2016).
5. B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* **361**, 360 (2018).
6. D. Weininger, *J. Chem. Inf. Model.* **28**, 31 (1988).
7. I. Anishchenko *et al.*, *Nature* **600**, 547 (2021).
8. M. Mardal *et al.*, *J. Anal. Toxicol.* **43**, 520 (2019).
9. M. A. Skinnider, R. G. Stacey, D. S. Wishart, L. J. Foster, *Nat. Mach. Intell.* **3**, 759 (2021).
10. E. J. Bjerrum, arXiv:1703.07076 [cs.LG] (2017).
11. M. A. Skinnider *et al.*, *Nat. Mach. Intell.* **3**, 973 (2021).

10.1126/science.adk8626

FINALIST

Isabella S. Bower

Isabella Bower received her undergraduate and postgraduate degrees in architecture from Deakin University and the University of Melbourne. After completing a PhD in psychology from Deakin University, she was awarded a prestigious postdoctoral research fellowship from the University of South Australia in 2023. Her research explores whether we can improve brain functioning and mental health through built environment design. **www.science.org/doi/10.1126/science.adk8629**

FINALIST

George Goshua

George Goshua received his MD from the University of Kentucky and completed his clinical residency and fellowship training in internal medicine and hematology-oncology at Yale University under the guidance of T. P. Duffy and A. I. Lee. He later completed an MSc focused on risk and decision science at the Harvard T.H. Chan School of Public Health under the supervision of A. Pandya. He opened his laboratory in the Department of Internal Medicine at the Yale School of Medicine in November 2022. His research applies health decision science methodology to identify equity-informed, cost-effective diagnostic and therapeutic treatment strategies for individuals living with hematologic and oncologic diseases.**www.science.org/doi/10.1126/science.adk8633**

PHOTOS: (TOP TO BOTTOM) COURTESY OF ISABELLA BOWER; JOEL CALLAWAY

Hallucinating hallucinogens

Michael A. Skinnider

Science **382** (6671), . DOI: 10.1126/science.adk8626