**Volume 41** Spring 2021

# MURJ

**Massachusetts Institute of Technology Undergraduate Research Journal**

Features p. 11

**Dog-Inspired Robot Noses**

News p. 6

**The Wax Propulsion Team @MIT Media Lab**


Develop your scientific career with our support
Search for your new role quickly by discipline, country, salary and more on go.nature.com/mit
nature careers
A101469

# Contents

**MURJ Staff**
MIT Undergraduate Research Journal

**Massachusetts Institute of Technology**

June 2021

Dear MIT Community,

We are thrilled to present the 41st issue of the MIT Undergraduate Research Journal. This semester has taken place during an unprecedented time, in which the MIT students, staff, and faculty have been physically separated to protect each other and our broader communities. The past year and a half have reinforced our belief in the power of scientific research and innovation, as well as the importance of community to science. We are proud and honored to continue in this spirit as we showcase the hard work and creativity of MIT students.

This issue features reporting on curriculum developments and fascinating research taking place on campus. Topics range from nature-inspired innovations for propulsion in space to the uniquely project-based NEET program for students to the development of cancer-sniffing robots inspired by dogs which can be trained to detect cancer with incredible accuracy. We additionally highlight a diverse array of original student research, featuring the development of a deep reinforcement learning framework with enhanced memory capabilities and an analysis of college reopenings for insight into broader COVID-19 policy decisions.

As always, we acknowledge that the biannual publication of this journal is the product of hard work, collaboration, and commitment by MURJ staff members and often a product of years of hard work and investment by undergraduate researchers and their mentors. We would like to thank our editorial board and contributors for their time and hard work this semester and for persevering through the challenges of a largely remote school year. In addition, we would like to thank all the undergraduates

**Massachusetts Institute of Technology**

**MURJ Staff**
**MIT Undergraduate Research Journal**

who shared their research with us and the greater MIT community.

For previous issues of the MIT Undergraduate Research Journal, please visit our website at murj.mit.edu. If you are interested in contributing to future issues of the MIT Undergraduate Research Journal, we would be delighted to have you. Please contact murj-officers@mit.edu if you have any questions or comments.

Best,

Natasha Joglekar
Co-Editor-in-Chief

Gabrielle Kaili-May Liu
Co-Editor-in-Chief

EDUCATION

# MIT New Engineering Education Transformation: Living Machines

*Students are able to tackle problems in biotechnology through teamwork, cutting-edge coursework, and critical-thinking workshops in project-based learning.*

**With the diverse set of skills MIT students learn** through different majors, the need for interdisciplinary teamwork is crucial for developing the latest technologies. The New Engineering Education Transformation (NEET) Living Machines thread exemplifies these principles through immersing students in biotechnology through diverse lenses. Much of the class focuses on humanizing drug development by creating 'organ-on-chip' technologies, where devices represent a specific organ or tissue in the human body that can be used for testing and development. Students of varying majors, including biological engineering, chemical engineering, computer science, mechanical engineering, and several others, are able to tackle problems in biotechnology through teamwork, cutting-edge coursework, and critical-thinking workshops in project-based learning.

When asked to speak about the intentions behind NEET Living Machines, instructor Dr. Medhi Salek commented, "The whole idea of NEET is project based learning. MIT and other universities have a lot of great courses that are offered for students, but there is still a part of the learning that needs to be done outside of the class. The whole NEET program is based upon project-based learning".

The courses required for NEET Living Machines differ from typical coursework in that they are primarily research based and cover more specific technical skills.

"The [Living Machines] thread offers a package for students. The interpersonal skills that students learn are really valuable for them. They learn how to do research projects in teams, which is useful especially in biotechnology where students need to know how to collaborate" added Dr. Salek, speaking about what benefits NEET Living Machines has for students.

The Living Machines thread has evolved over time, originally co-founded by Professor Linda Griffith of the Biological Engineering department. The structure of NEET Living Machines is track-based, where there are five different tracks for students to get an additional focus within biotech--synthetic biology, tissue engineering, computational biology, immunology, and microfluidics.

NEET Living Machines student Julia Van Cleef is a sophomore planning on working with the computational biology track. She spoke about the most valuable part of the program thus far, adding:

> "Based on my track choice and other research experiences, I would normally never have the opportunity to work on projects relating to microfluidics. However, this semester we were able to go through the entire process of designing microfluidic devices from using CAD and COMSOL softwares to model the device to actually going into the lab to 3d print molds and create a functional device. It was great to experience this process from start to finish and gain exposure to a relevant field in biotech."

Due to COVID-19, the time spent in the lab for the NEET Living Machines classes has been limited, so the instructors are hopeful that the future would allow for more lab work.

When asked about the next steps for the program, Dr. Salek mentioned Living Machines would like to add even more interdisciplinary team projects--specifically having students work on a more independent research project in the lab. This would allow for students to complete the program research requirements, while consequently emphasizing the unique aspects of NEET Living Machines.

The breadth of programs and academic diversity at MIT is one of the many reasons why it is such a special institution. Programs like NEET Living Machines exemplify this, allowing for students to enhance their time spent at MIT by gaining invaluable skills for the future.

— Tatum Wilhelm

AEROSPACE

# Space Enabled Research Group's Wax Propulsion Team at MIT Media Lab

*Revolutionizing the equity, safety, and affordability of in-space propulsion*

**In-space, propulsion is truly a wonder.** To be able to move a satellite in-space in a certain direction despite its orbital motion, is an interesting endeavor, to say the least. Many different propulsion systems can be used to accomplish this, but the MIT Media Lab Space Enabled Research Group's Wax Propulsion team is focusing on a green, non toxic solution for in-space propulsion that can also increase equity and justice in-space technologies. Their solution is the use of paraffin and beeswax, which are wax-like substances. These materials are low cost, safe, and are already used and located in mainstream satellites because they are used in wax thermal insulation. Past studies at Stanford University and the University of Tennessee have proven that paraffin and beeswax can work as fuel grains for propulsion.

The Wax Propulsion team aims to find a way to reheat the wax used in thermal insulation to form a solid fuel grain cell. The fuel grain cells for propulsion must be in the shape of an annulus. An annulus is a hollow cylinder shape. In order to produce this in a space or microgravity environment, the best option would be to spin melted wax in a cylinder tube. This would allow the centrifugal force produced by the spinning to push the melted wax to the surface of the cylinder, producing an annulus.

The Wax Propulsion team is hoping to analyze the rates of annulus formation in 1g and microgravity environments, so they can make accurate predictions about how difficult it would be to create fuel grains made of paraffin and beeswax in-space.

There are two steps to understanding annulus formation in-space: creating an annulus shape, and understanding how long it would take for the wax to solidify in the annulus formation. As a result, the Wax Propulsion team has organized a five step process that delves into these important aspects of the project. First, they conducted an experiment that tested the solidification process for paraffin wax in a laboratory setting (1g environment). Then, they conducted an experiment in the laboratory that determined the best rotational speeds and casting times, which would be the amount of time that complete annulus formation would take. Next, there would be two microgravity flights: one to test a liquid water-filled centrifuge that will allow them to examine the fluid mechanics of the system, and another to determine the rotation rates at which annulus formation occurs, to compare to laboratory results. Lastly, there will be an experiment that uses paraffin to understand the thermodynamics and fluid mechanics of centrifugal casting in longer time intervals in microgravity environments.

*"Paraffin and beeswax can work as fuel grains for propulsion."*

In my UROP this spring, I had the opportunity to work with this team and learn more about propulsion and

*SpaceX's Dragon spacecraft lifts off from Cape Canaveral Air Force Station, Florida.*

experimenting in microgravity environments. I helped create drawings of parts of the centrifugal casting system that could be used to send to machinists so they understand how to produce each part. I also learned about the image analysis process for this project, which is integral for collecting and recording data for the solidification of paraffin and the optimal rotation speeds and times for annulus formation.

This summer, I will be continuing my work with the Space Enabled Wax Propulsion team to apply my knowledge of image analysis. I will be collecting and recording data from a recent microgravity flight. This data will be used to examine what the optimal rotation rates of paraffin, oil, and water are in microgravity environments, which will be compared to our ground based experiments.

Overall, the work at Space Enabled involves the comparison of materials in microgravity environments and laboratory environments. Their work will culminate in a rigorous analysis of the ability of wax propellants to be used for in-space propulsion, which could revolutionize the equity, safety, and affordability of in-space propulsion.

— Dinuri Rupasinghe

gsk do more feel better live longer

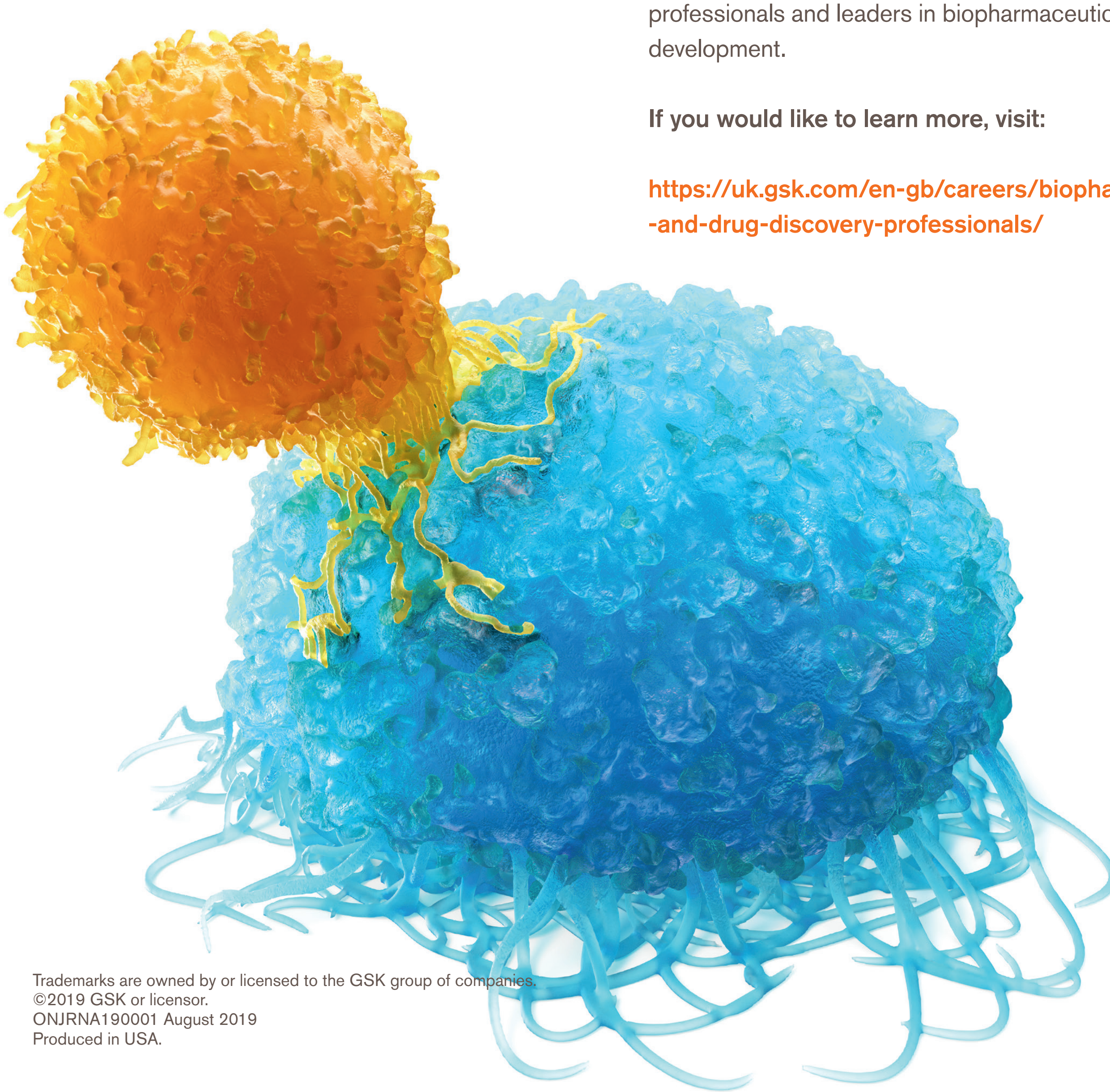
# A healthcare company with a special purpose

Our goal is to be one of the world's most innovative, best performing and trusted healthcare companies. We believe that Biopharm medicines are key to achieving our ambitions in immuno-oncology and other immunologically driven diseases. As a global enterprise with activities in the US and the UK, we have a thriving team of Biopharmaceutical and Data Scientists across discovery, development and manufacturing.

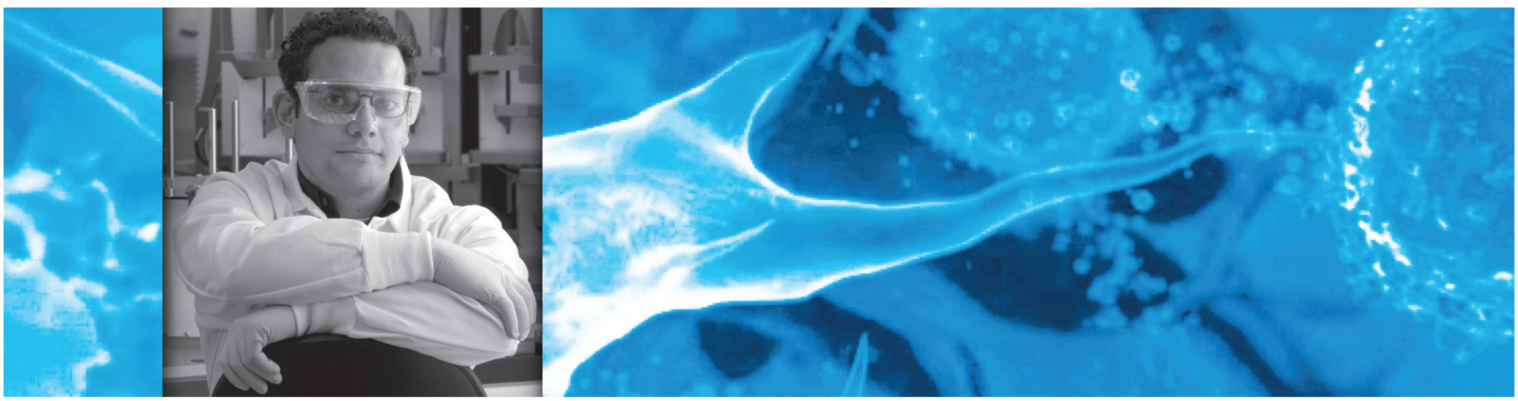
We are growing our team to deliver our portfolio of exciting Biopharm medicines to patients around the world and have opportunities available across all disciplines for junior scientists, experienced professionals and leaders in biopharmaceutical development.

**If you would like to learn more, visit:**

https://uk.gsk.com/en-gb/careers/biophar
-and-drug-discovery-professionals/



T-cell attacking cancer tumour



# MURJ Features

# Dog-inspired robot noses: using smell technology to detect prostate cancer

By Celina Zhao

***"The hospital and our best technology were wrong, but the dog was right!" exclaims Dr. Andreas Mershin, researcher and inventor at MIT's Center for Bits and Atoms.***

He's referring to a 2004 study[1] that trained dogs to detect bladder cancer from samples of urine. When sniffing the urine sample of one healthy, cancer-negative man, one dog continually insisted that the sample was positive, despite all hospital records and other diagnostic tests indicating otherwise. At the time, it was brushed off as a mistake: a false positive. But months later, the person was diagnosed with early-stage bladder cancer. The dog had detected the cancer not only accurately, but much, much earlier than any human-built system could.

Known for centuries as man's best friend, dogs offer much more than just cuddles and companionship. Numerous studies have demonstrated that trained dogs can sniff out many kinds of diseases, including several kinds of cancer and Covid-19. In fact, dogs have even been able to identify positive prostate cancer samples with 99% accuracy.[2]

Inspired by these canine olfactory abilities, Dr. Mershin and researchers from Johns Hopkins University School of Medicine, Prostate Cancer Foundation, and the nonprofit organization Medical Detection Dogs in the UK formed a team to integrate several prostate cancer detection methods.

The group aimed to explore the feasibility of combining trained dog detection, lab-based tests, and machine learning (through artificial neural networks trained to recognize patterns in the data from the dogs and lab tests), with a final goal of building a new mechanical odor-detection system compact enough to fit into a cellphone. Their findings were published in February 2021 in the journal *PLOS One*.[3]

***"The dog had detected the cancer not only accurately, but much, much earlier than any human-built system"***

Prostate cancer is the second leading cause of cancer for men in the developed world. However, current detection methods are acutely lacking. The primary diagnostic test, called a PSA test, screens the patient's blood levels for elevated levels of a protein called prostate- specific antigen. Not only has the inventor of the test himself described it as hardly better than a coin toss,[4] but the follow-up steps are also less than ideal.

Prostate cancer comes in two primary groups: five fast-moving, lethal types that require aggressive treatment, and twelve slow-moving types that will turn malignant if disturbed or

1 https://www.bmj.com/content/329/7468/712

2 https://www.health.harvard.edu/blog/trained-dogs-can-sniff-prostate-cancer-201406302101

3 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245530

4 https://www.nytimes.com/2010/03/10/opinion/10Ablin.html

treated aggressively. To distinguish which category the tumor is, an invasive biopsy requires a large needle to be jabbed through the walls of the rectum in order to retrieve a tissue sample from the prostate. In a somewhat Catch-22 moment, trying to diagnose which type of prostate cancer is present might result in disturbing a previously-benign tumor enough to make it deadly.

***"There's virtually no way for current laboratory tests alone to imitate the canine intuition"***

This is where smell comes in. Our canine companions have incredible ability to diagnose prostate cancer at extremely high accuracy from just a few seconds of sniffing. If we could automate these olfactory abilities in a machine, the process of prostate cancer diagnosis would become significantly less harrowing. The challenge? Even after decades of research, scientists still have little to no understanding of how our sense of smell works.

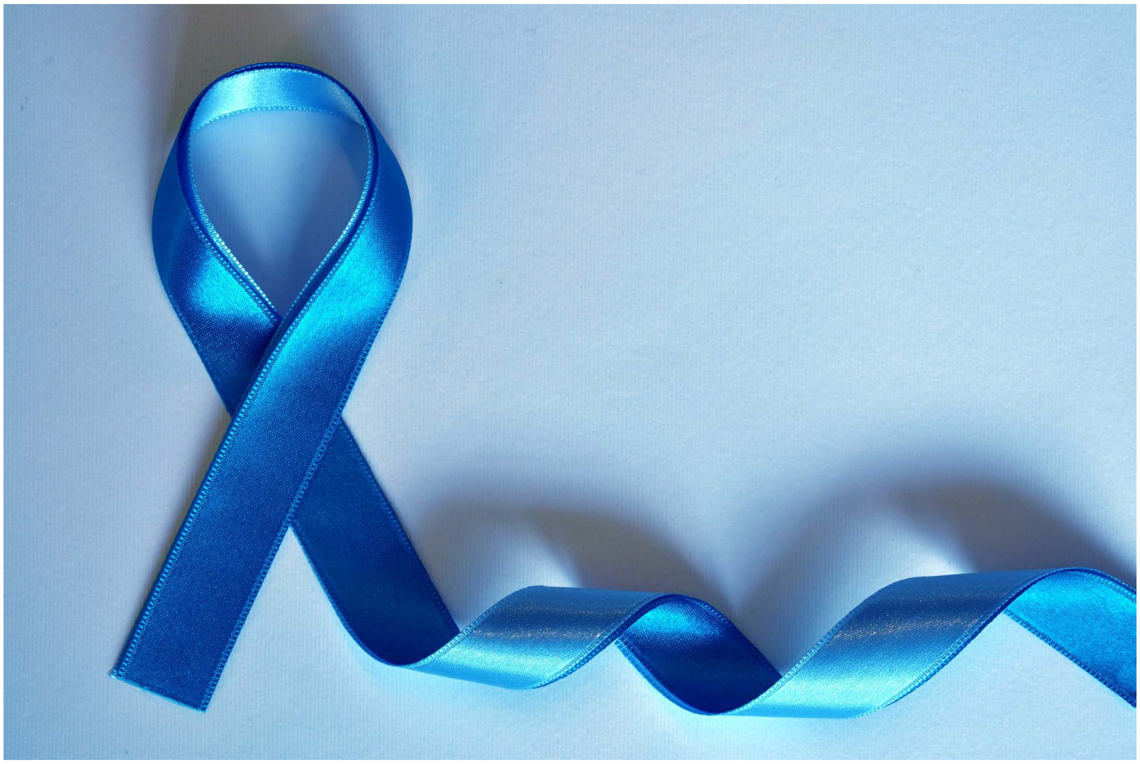
"If I give you the ingredients of a cake – eggs, flour, sugar – you don't know what the cake is going to taste like," Dr. Mershin explains. "The tasting part happens in the mouth. The scent character emerges out of the molecules, but it's a function of the nose and the brain and the prior training you've done that determines what experience you're going to have." Essentially– what something is made of is not the same as what it smells of.

On the other hand, analytical techniques within the lab depend on having a list of molecules – biomarkers – by name and concentration. As a result, there's virtually no way for current laboratory tests alone to imitate the canine intuition.

An integrated approach with machine learning, however, is a viable solution. In humans and dogs, the brain is responsible for computing the origin and composition of smell. Dr. Shuguang Zhang, a biochemist at MIT Media Lab's Laboratory for Molecular Architecture who was not involved in the study, further describes the pattern recognition process as a "QR code in 3-D," where the compares the code to a huge database of smells it has come in contact with in the past.

To begin exploring the possibility of uniting machine learning with dog detection and lab tests, the study used 50 urine samples from Johns Hopkins University Hospital: 12 men with biopsy- confirmed Gleason 9 (high-grade and advanced) prostate cancer and 38 men with negative biopsies. First, the trained dogs were able to identity samples with 71% sensitivity (detecting truly positive cases) and 70-76% specificity (detecting truly negative cases).

Next, the samples were processed through gas chromatography-mass spectroscopy to identity the individual vapor molecules within the samples, as well as microbially profiled to break down the genetic composition of microbial species in the urine. Given these two data sets, artificial neural networks were trained to identify specific portions of the spectroscopy and profiling data that factored into the dogs' diagnoses and



*A light blue ribbon is generally used to support those living with prostate cancer and promote awareness.*

specific differences between positive and negative samples.

"For any artificial intelligence design, you need a huge sample size," Dr. Zhang says of the study. "For scientists to continue developing the tools needed for automating smell, we need to build a much, much larger database."

Morgan Moncada, Director of Product and Operations at the Silicon Valley-based company Aromyx, a start-up aiming to quantify taste and smell through biosensor technology, says, "The study is preliminary, but exciting. The data is showing that there is potential in this space."

Though both Dr. Zhang and Moncada agree that there is much more work to be done, the possibilities are still thrilling. Dr. Mershin, for one, is very optimistic that smartphones with smell technology will be a huge revolution in the medical stream. "Birds taught us how to fly, but we don't bind a bunch of birds to an airplane," Dr. Mershin says. "Similarly, dogs are teaching us to sense all these diseases, and the solution is to put this new sense into technology that already has vision, geolocation, and audio. Having all of these data flow together will transform society and the way we approach disease detection."

# MURJ

# UROP Summaries

# Analyzing Public Health through College Reopening Decisions

**Julian Zulueta[1], Denise Koo[2]**

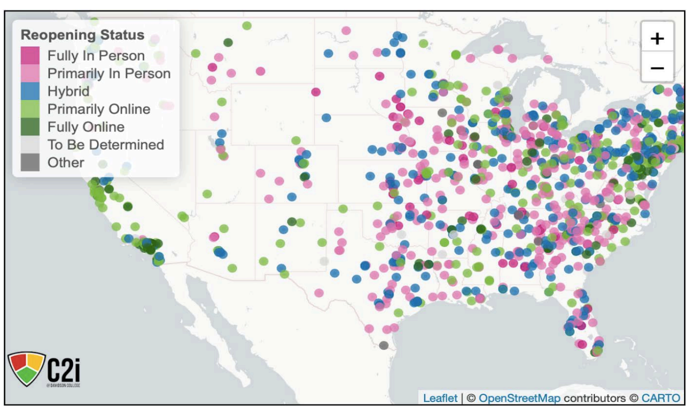
1 Student Contributor, MIT Department of Biological Engineering, Cambridge, MA 02139

2 Supervisor, Centers for Disease Control and Prevention Foundation, Atlanta, GA 30308

## 1. Introduction

The transmission of COVID-19 in the United States has altered the college dynamic in 2020. Here, higher-ed institutions have had nuanced decisions on whether to transition to online mediums, such as Zoom Video Communications, or remain in-person during the fall.

In order to better comprehend different colleges' decisions, states' COVID-19 cases, annual GDP, and political party affiliation were analyzed to determine the significance of these factors' contributions to college fall plans. Each factor is analyzed independently to see how it affects public decisions, such as reopening colleges. These correlations can help public health officials understand how regions respond to COVID-19 and utilize the information to adopt necessary health precautions.




College Crisis Initiative by Davidson College

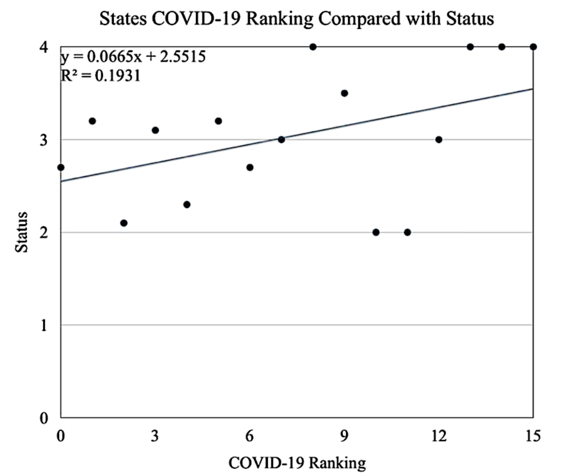
## 2. Results & Discussion

The visual data for this study is located in Figures 1-3, where the three external factors are compared to approximately one-third of the United States. The data provides the slopes and r-squared data, which were calculated to determine the prominence and accuracy of this information. In Figure 1, "COVID-19 Cases Compared to States' Decisions," data is based from the CDC COVID Data Tracker on July 30, 2020 at 5:45 PM. Here, the top eight states and bottom eight states are numbered on a scale 0-15. This exemplifies a gradient to display states with higher rates of COVID-19 closer to 0, while states with lower rates are closer to 15. Additionally, each state is given a numerical status score that ranges from 1-5. The score is based on the College Crisis Initiative, where each college decision is given a number. To illustrate, "Fully Online" would constitute a 1, whereas "Fully in Person" would receive a 5. Numbers in between this range would be representative of other possible decisions—Primarily Online (2), Hybrid (3), and Primarily in Person (4). Schools that were listed as "TBD" or "Other" were not given a score. The top ten schools for determining each state's score were chosen based on the 2020 Best National University Rankings in the U.S. News & World Report to make this assessment. This data was inputted on a scatter plot and lines of best fit were created to determine a slope and r-squared factor.

Figures 2-3 adhered to similar procedures. Figure 2 is based on the U.S. Bureau of Economic Analysis (BEA), displaying states with a higher GDP closer to 0 on the x-axis. Figure 3 is based on the Gallup 2017 U.S. Party Affiliation by State, which demonstrates the percentage of Republican or Democratic lean. States with a greater percentage of a Republican lean are closer to 0 on the x-axis, whereas Democratic leaning states are closer to 15. By utilizing updated sources, the information on the figures best displays an accurate portrayal of how the factors individually affect college fall decisions to reopen.

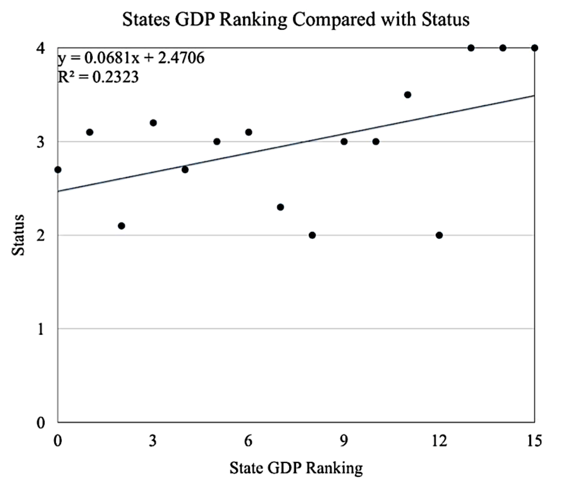
The study reveals that r-squared factors are low, and only Figure 3 is statistically significant. This indicates that no single social factor in isolation has a major impact in determining public health.

Nevertheless, primary constraints that this study had include the sample size of the data and the lack of qualitative elements, such as a region's culture in determining public health decisions. A broader dataset and further analysis, including modeling the interactions between the three factors, would be helpful in determining the significance and strength of external components on public health.



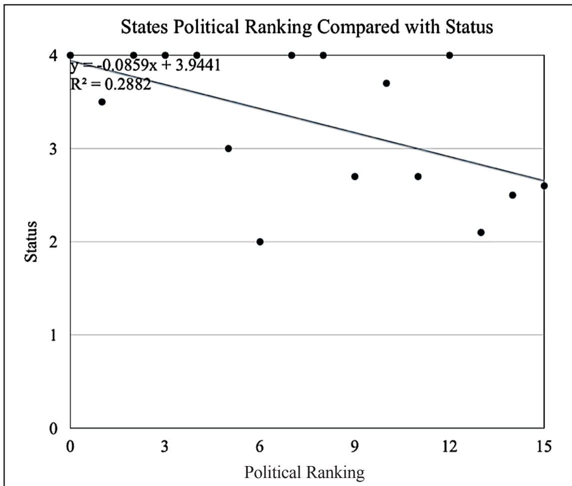

| Ranking | State | Status |
|---|---|---|
| 0 | California | 2.7 |
| 1 | Florida | 3.2 |
| 2 | New York | 2.1 |
| 3 | Texas | 3.1 |
| 4 | New Jersey | 2.3 |
| 5 | Georgia | 3.2 |
| 6 | Illinois | 2.7 |
| 7 | Arizona | 3.0 |
| 8 | West Virginia | 4.0 |
| 9 | North Dakota | 3.5 |
| 10 | Maine | 2.0 |
| 11 | Montana | 2.0 |
| 12 | Alaska | 3.0 |
| 13 | Wyoming | 4.0 |
| 14 | Hawaii | 4.0 |
| 15 | Vermont | 4.0 |

**Figure 1.** COVID-19 Ranking Compared with Status




| Ranking | State | Status |
|---|---|---|
| 0 | California | 2.7 |
| 1 | Texas | 3.1 |
| 2 | New York | 2.1 |
| 3 | Florida | 3.2 |
| 4 | Illinois | 2.7 |
| 5 | Pennsylvania | 3.0 |
| 6 | Ohio | 3.1 |
| 7 | New Jersey | 2.3 |
| 8 | Maine | 2.0 |
| 9 | Rhode Island | 3.0 |
| 10 | Alaska | 3.0 |
| 11 | North Dakota | 3.5 |
| 12 | Montana | 2.0 |
| 13 | South Dakota | 4.0 |
| 14 | Wyoming | 4.0 |
| 15 | Vermont | 4.0 |

**Figure 2.** States' GDP Ranking Compared with Status




| Ranking | State | Status |
|---|---|---|
| 0 | Utah | 4.0 |
| 1 | North Dakota | 3.5 |
| 2 | Wyoming | 4.0 |
| 3 | Idaho | 4.0 |
| 4 | South Dakota | 4.0 |
| 5 | Alaska | 3.0 |
| 6 | Montana | 2.0 |
| 7 | Nebraska | 4.0 |
| 8 | Hawaii | 4.0 |
| 9 | Illinois | 2.7 |
| 10 | Connecticut | 3.7 |
| 11 | California | 2.7 |
| 12 | Vermont | 4.0 |
| 13 | New York | 2.1 |
| 14 | Maryland | 2.5 |
| 15 | Massachusetts | 2.6 |

**Figure 3.** States' Political Ranking Compared with Status

# MURJ Reports

# Dreaming with Transformers

***Jordan Docter[1], *Catherine Zeng[2], Alexander Amini[1], Igor Gilitschenski[3], Ramin Hasani[1], Daniela Rus[1]**

1 Computer Science & Artificial Intelligence Laboratory, MIT

2 Harvard University

3 Toyota Research

** indicates authors with equal contributions*

**Deep models built by self-attention mechanisms (e.g., Transformers [1]) proved to be effective in most application domains ranging from natural language processing to vision. This is a direct result of their ability to perform credit assignment in long time horizon and their scalability with massive amount of data. In this paper, we explore the effectiveness of the Transformer architecture in model-based deep reinforcement learning (RL). The generalization performance of a model-based deep RL agent highly depends on the quality of its state transition model and the imagination horizon. We show that Transformers improve the performance of model-based deep RL agents in environments which require long time horizon. To this end, we design the *Dreaming with Transformers* experimental framework which learns deep RL policies by using Transformers in state-of-the-art model- based RL frameworks (e.g., Dreamer [2, 3]). Our preliminary results on the multi- task DMLab-30 benchmark suite suggest that Dreaming Transformers outperform RNN-based models in challenging memory environments.**

## 1. Introduction

In this work, we present Dreaming with Transformers, a deep reinforcement learning (RL) framework that captures the temporal dependency of environment interactions to construct world models and optimal policies within challenging memory environments. Deep RL tasks cover a wide range of possible applications with the potential to impact domains such as robotics, healthcare, smart grids, finance, and self-driving cars. The fundamental challenge in these domains is how to perform credit assignment in long time horizon tasks, and how to obtain deep RL policies that can be transferred to the real world.

Recent model-based RL frameworks [2, 4] suggested improvements over world models [5] that explicitly represent an agent's knowledge about its environment. World models facilitate generaliza- tion and can predict the outcomes of potential actions (in an imagination space) to enable decision making. It has been shown that they achieve state-of-the-art performance across a series of standard RL benchmarks [3]. This model is called *Dreamer* which presents agents that can learn long-horizon behavior directly from high-dimensional inputs, solely by latent imagination.

Dreamer agents use an actor-critic algorithm to compute rewards and uses recurrent neural networks (RNNs) to make predictions within a latent imagination state-space. The use of RNNs and their gated versions such as the long short-term memory (LSTMs) [6] and Gated recurrent units (GRU) [7] is natural due to the spatiotemporal characteristics of the challenging memory environments. However, the memory span of a recurrent network is limited [8], their information processing mode is sequential and mutual information of RNNs decay exponentially in temporal distance of sequence inputs [9].

Recently, deep architectures based on attention mechanism proved to enable parallel credit assignment in very long sequences, and significantly outperformed recurrent models [1]. These architectures which are called Transformers are the primary choice in natural language processing (NLP) tasks [10], and are becoming the dominant architecture in vision tasks as well [11]. Recent works in model-free RL provided additional evidence that Transformers can be effective in challenging memory tasks in the context of RL [12]. In the present study, we claim that Transformers [1] are more stable and performant in capturing mutual information over long time horizons, and are thus better candidates for learning world models.

In particular, we improve upon model-based RL frameworks by specifically addressing an agent's capacity for world representations and credit assignment in long-horizon tasks. We propose an architecture that marries the concept of self-attention seen in transformers into an agents representation of the world. Our main contribution lies in the architecture of the latent dynamics model which particularly allows for non-markovian transitions of the latent space. This freedom greatly increases the predictive power of imagined trajectories, which in turn yields more optimal actions. In section 2, we provide background on the architecture of world models as well as the concept of self-attention. In section 3, we provide an overview of our proposed architecture. We motivate our choices by providing analysis of the shortcomings of current models specifically within the context of highly complex environments. In section 4, we give our experimental setup and present preliminary results on a set of complex benchmark RL tasks. In section 5 we discuss future work and goals of this architecture, namely related to real world applications.

## 2. Problem Setup

### 2.1 Reinforcement Learning

The objective of reinforcement learning is to search for an optimal policy in a Partially Observable Markov Decision Process (POMDP) defined as $(\mathcal{S},\mathcal{A},\mathcal{P},\mathcal{R},\mathcal{O})$ [4]. Particularly in a *partially* observable MDP, the agent makes observations of the environment that may only contain partial information about the underlying state. Formally, we let $o_t$, $r_t$, $s_t$, $a_t$ be the observation, reward, state, and action at time step $t$ in $\{1, \ldots, T\}$. The at each time step $t$, the agent will generate and execute an action $a_t \sim p(a_t \mid o_{\leq t}, a_{\leq t})$. The environment will change to a new state according to some transition probability function $s_t \sim P(s_t \mid s_{t-1}, a_t)$, but the agent will only receive observations and reward $o_t$, $r_t \sim p(o_t, r_t \mid o_{\leq t}, a_{\leq t})$ from the environment. The goal of the agents is to maximize the expected reward $\mathbb{E}(\sum_{t=0}^{T} r_t)$.

### 2.2 Model-Free vs Model-Based RL

Recent reinforcement learning models [2, 4] have found success by learning world models that explicitly represent an agent's knowledge about its environment. World models stand in contrast to model-free frameworks which directly learn a correspondence between the state-space and action- space. It is shown [13] that in large unknown environments, model-free frameworks suffer from low sample efficiency and high sample complexity, and in some cases are not optimal. World models attempt to address this issue by providing the means for agents to extrapolate in situations they have never encountered before. This is accomplished by learning a representation of the world in a latent space, and then forming policies on top of this latent space.

## 3. Dreaming with Transformers

We consider reinforcement learning tasks with highly complex state and action spaces such as images and continuous movement within the environment. Inspired by recent works in model-based RL and sequence-to-sequence machine learning models, we propose a deep reinforcement learning model with two key components: a world representation (3.1), and latent imagination with transformers (3.2). Our main contribution lies in the marriage of transformers into world models.

### 3.1 World Model

Our world model consists of several high level components: (1) an encoder from observations (images) to a latent state space, (2) a latent dynamics model that imagines trajectories in the latent space, and (3) an actor-critic model that predict actions and reward of imagined trajectories. The agent makes decisions by imagining trajectories in the latent space of the world model based on past experience, and estimating trajectory rewards through learned action and value models. In this work, we focus on the latent dynamics component. We first define the following:

- $o_t$ is the observation at time $t$
- $\hat{o}_t$ is the reconstructed observation at time $t$
- $a_t$ is the action at time $t$
- $s_t$ is a stochastic state at time $t$ that incorporates information about $o_t$
- $\hat{s}_t$ is a stochastic state at time $t$ that does not incorporate information about $o_t$
- $h_t$ is the deterministic state from which the $s_t$ and $\hat{s}_t$ are predicted off of
- $\mathcal{M}$ is the memory length of the sequential model.

The model can thus be formulated by the following distributions where we use $p$ for distributions that generate samples in the real environment, $q$ for their approximations that enable latent imagination and $\phi$ to describe their shared parameters:

- Transformer model: $h_t \sim f_\phi(h_{[t-\mathcal{M}, t-1]}, s_{[t-\mathcal{M}, t-1]}, a_t)$
- Representation model: $s_t \sim p_\phi(h_t, o_t)$
- Transition model: $\hat{s}_t \sim q_\phi(h_t)$
- Image model: $\hat{o}_t \sim q_\phi(h_{t-1}, s_{t-1})$
- Reward model: $r_t \sim q_\phi(h_t, s_t)$

The representation model encodes observations and actions to create continuous states st with non-markovian transitions. The transition model predicts future states in the latent space without seeing the corresponding observations that will later cause them. The image model reconstructs observations from model states. The reward model predicts the rewards given the model states. The policy is formed by imagining hypothetical trajectories in the compact latent space of the world model using the transition model, and choosing actions that maximize expected value.

We refer the reader to Hafner et al.'s [2] work for a more detailed description of the remaining components of the architecture which we largely base ours off of.

### 3.2 Reinforcement Learning

Our model imagines trajectories in the latent space via transformers. Transformers [1] are neural nets that *transform* a given sequence of elements, such as the sequence of words in a sentence, into another sequence. Similarly to other sequence-to-sequence architectures, they consist of encoders and decoders to produce an output sequence from an input sequence. Recent works have shown that Transformers achieve staggering improvement over previous sequence-to-sequence models. Their key advantage lies in *attention*. For each input that the (for example) LSTM reads, the attention-mechanism takes into account several other inputs at the same time and decides which ones are important by attributing different weights to those inputs.

By analyzing the auto-mutual information (across time lags) of sequence-to-sequence models, Shen [9] shows that the mutual information decays exponentially in temporal distance in RNNs, whereas, long-range dependence can be captured efficiently by Transformers. The sequential data within sophisticated reinforcement learning tasks, such as self-driving cars, are highly correlated across time. As such, it is natural that Transformers have potential to better represent the latent state space and make predictions of future states.

As shown in figure 1, the transformer takes the past $M$ deterministic states $h_{[t-M,\, t-1]}$, stochastic states $s_{[t-M,\, t-1]}$, and action at to predict future states $h_t$. Observations are encoded via an encoder/decoder model. The transformer imagines future states $h_{\geq t}$ off of past $h_{[t-M,\, t-1]}$ and $s_{[t-M,\, t-1]}$, and the most recent action $a_{t-1}$. The imagined states are used to imagine the world (states, value, reward) in the future $\hat{s}_{\geq t}$, $\hat{v}_{\geq t}$, $\hat{\boldsymbol{r}}_{\geq t}$, and find optimal policies $\hat{a}_{\geq t}$ within the imagined space. The hat operator (ˆ) indicates values that are predicted without their corresponding observations.

## 4. Preliminary Experiments and Results

Our preliminary experiments test the performance of our model on four environ- ments in the DMLab domain [14]: `rooms_collect_good_objects`, `rooms_watermaze`, `explore_object_rewards_few`, and `explore_obstructed_goals_large`. As a baseline, we compare against the Dreamer model, which uses a Gated Recurrent Unit (GRU) instead of a trans-former for imagination.

For our tensors to not exceed available memory, we ran these preliminary experiments with batch size 20, which is less than the batch size of the Dreamer model (50). We also compared the performance of our model with a naive transformer architecture and its performance with the stable transformer architecture as described in Parisotto et al [12]. In our experiments, we

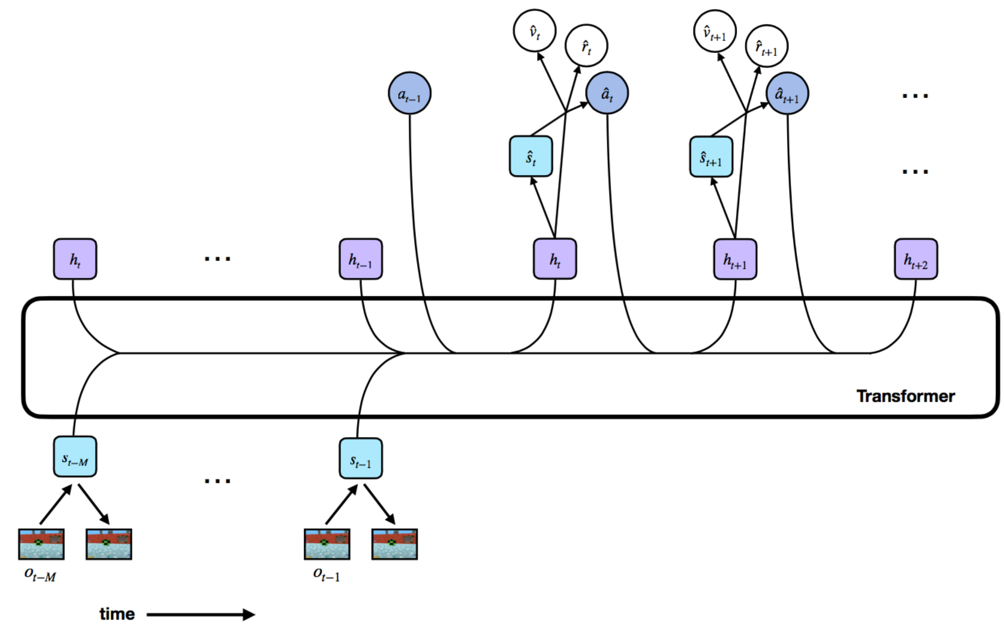



**Figure 1**: Latent imagination by transformer.

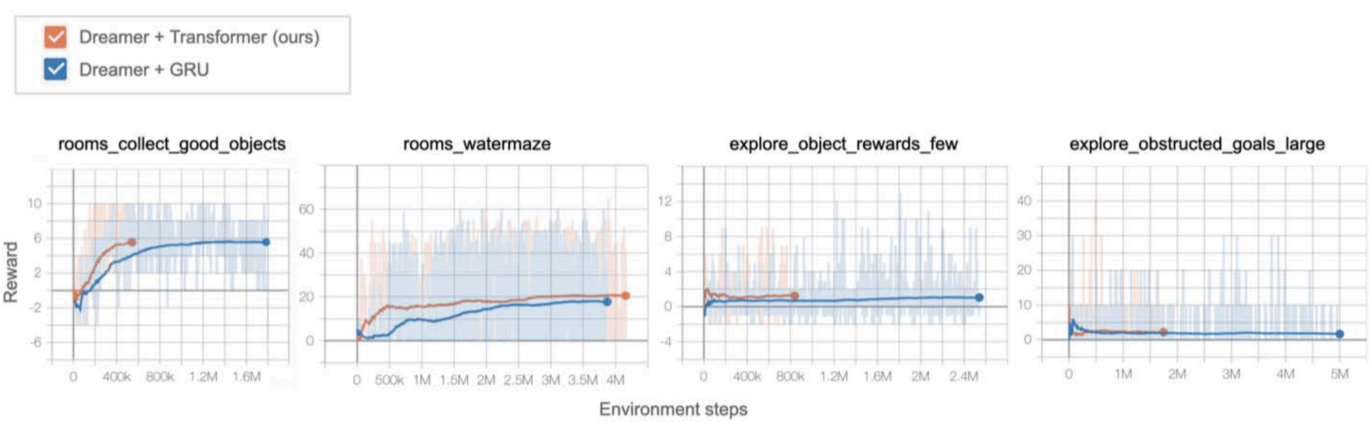



**Figure 2**: Preliminary results: Tensorboard graphs comparing the performance of our model to that of the original Dreamer agent (Dreamer + GRU). Faded lines repre- sent unsmoothed values, and darkened lines represent fully smoothed values. With minimal tuning and smaller batch sizes, our model performs better than Dreamer + GRU on `rooms_collect_good_objects` and `rooms_watermaze`, comparably to `explore_object_rewards_few` and `explore_obstructed_goals_large`.

found the stable transformer architecture capable of consistently outperforming the naive transformer on the DMLab tasks, so we used the stable transformer architecture in our model. We also took several steps towards hyperparameter tuning, including varying the transformer memory length, Dreamer's horizon length, agent exploration amount, and Dreamer's actor model entropy.

Despite the smaller batch size and minimal hyperparameter tuning, our model shows improvement over the original Dreamer model. Figure 2, shows a comparison of our model and Dreamer for a selection of reinforcement learning tasks (`rooms_collect_good_objects`, `rooms_watermaze`, `explore_object_rewards_few`, and `explore_obstructed_goals_large`).

## 5. Future Work

Ultimately, Dreaming with Transformers aims to address real world complex tasks where the action space is continuous and the observations are high dimensional. We are currently testing our model on VISTA [15], a data driven simulation for autonomous driving. VISTA evaluates an RL agent's ability to train by driving along roads in simulation and deploy into the real world.

Some additional directions of work include further hyperparameter tuning and investigating possible modifications to the Dreamer and stable transformer architectures. Since our work involves a large model-based RL agent in a complex setting, we anticipate running into complexities that have not been encountered in previous works.

## References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[2] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination, 2020.

[3] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models, 2020.

[4] Jingbin Liu, Xinyang Gu, and Shuai Liu. Reinforcement learning with world model, 2020.

[5] David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 31. Curran Associates, Inc., 2018.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[7] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259, 2014.

[8] Mathias Lechner and Ramin Hasani. Learning long-term dependencies in irregularly-sampled time series. arXiv preprint arXiv:2006.04418, 2020.

[9] Huitao Shen. Mutual information scaling and expressive power of sequence models, 2019.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

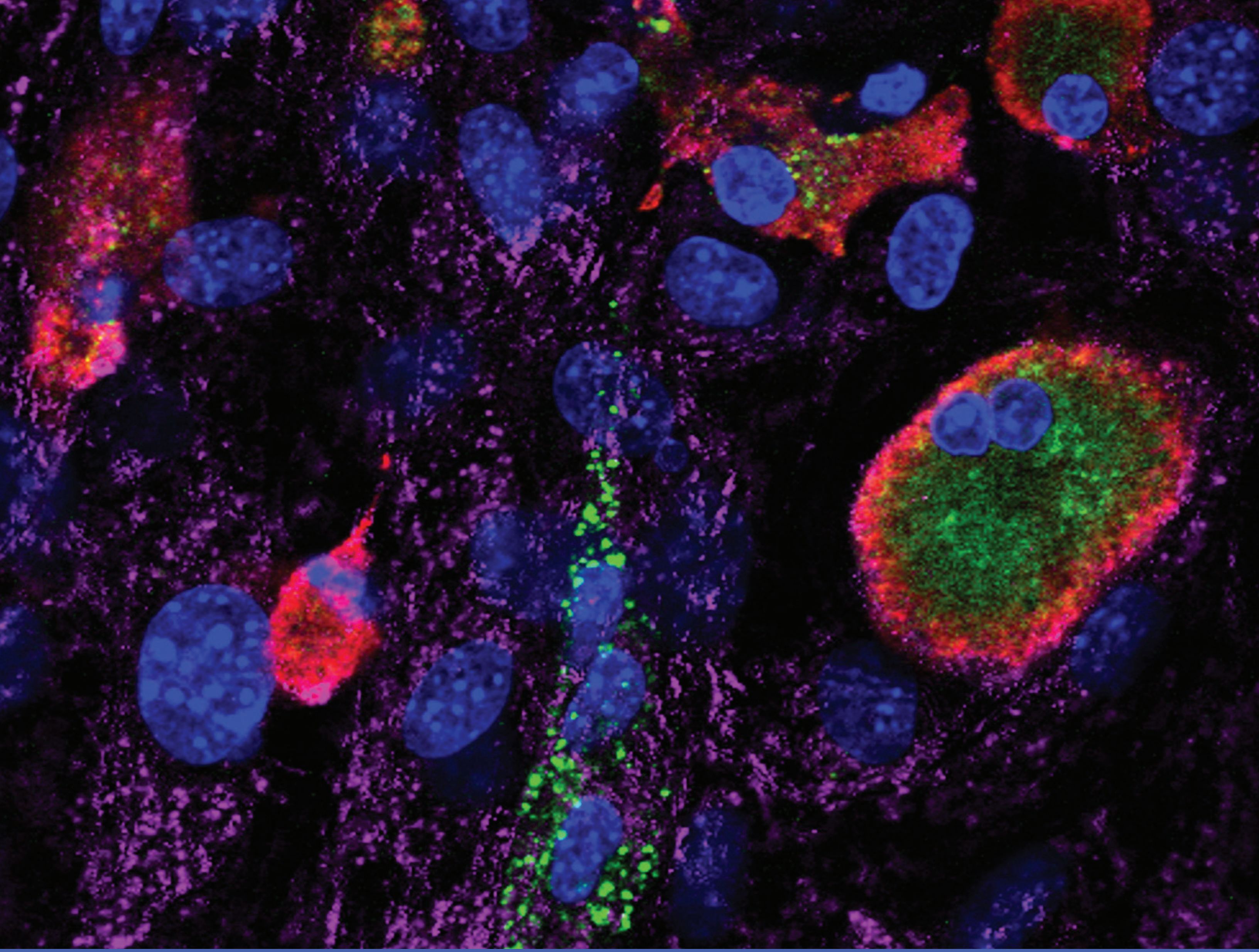
[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.

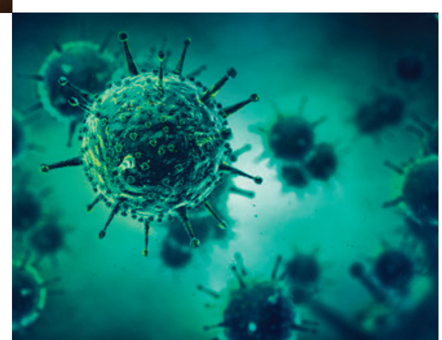
[12] Emilio Parisotto, H. Francis Song, Jack W. Rae, Razvan Pascanu, Caglar Gulcehre, Siddhant M. Jayakumar, Max Jaderberg, Raphael Lopez Kaufman, Aidan Clark, Seb Noury, Matthew M. Botvinick, Nicolas Heess, and Raia Hadsell. Stabilizing transformers for reinforcement learning, 2019.

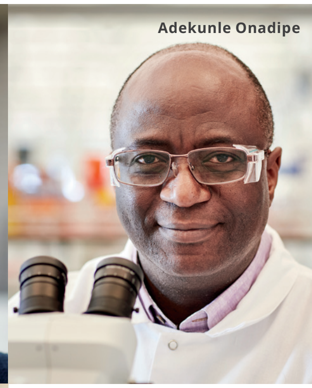
[13] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018.

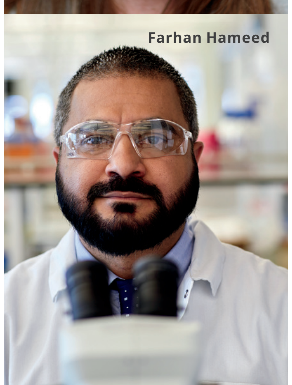
[14] Charles Beattie et al. Deepmind lab, 2016.

[15] Alexander Amini, Igor Gilitschenski, Jacob Phillips, Julia Moseyko, Rohan Banerjee, Sertac Karaman, and Daniela Rus. Learning robust control policies for end-to-end autonomous driving from data-driven simulation, 2020.










Everything that makes us unique makes us uniquely good at the work we do together.

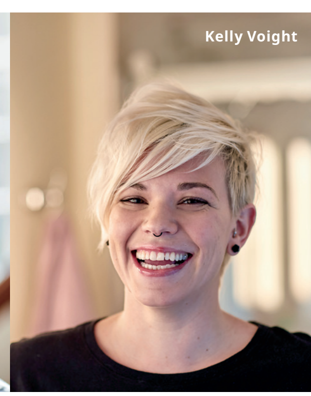
WE ARE THE MANY, DARING, DIFFERENT PEOPLE OF PFIZER –

WORKING TOGETHER TO DELIVER **BREAKTHROUGHS THAT CHANGE PATIENTS' LIVES.**

Pfizer

Discover what we're all about at

**pfizer.com/careers**



PALL CORPORATION

# Looking for a job that makes a difference?

**At Pall Corporation, we are unified by a singular drive: to take on our customers' biggest challenges and deliver complete solutions.**

To resolve the critical problems that stand in the way of achieving their goals.

**To help safeguard health.**

Pall's technologies play key roles in the development and manufacture of life-saving drugs that range from Ebola and COVID-19 vaccines to cancer-curing monoclonal antibodies. Our portfolio helps customers bring vital drugs to market faster at lower costs.

Are you ready to join a team that is making an impact in lives around the world every day?

Our company offers career opportunities at every level, giving you the chance to join our globally diverse team and apply your talents to meaningful work.

R&D — Process Engineering — Operations
Sales — Marketing — and more!

**No matter what, no matter where, we innovate and collaborate to deliver the one thing our customers need most:**

The unsolvable, **solved.**


**If you are looking for a job that makes a difference and gives you a true purpose each day, join our biotech team.**

**Discover the opportunities waiting for you here!**

**Visit http://bit. ly/PALLJobs**

**Scan me**