Received: 5 July 2023 | Accepted: 6 October 2023

DOI: 10.1002/aaai.12163

SPECIAL TOPIC ARTICLE

Institute for Foundations of Machine Learning (IFML): Advancing AI systems that will transform our world

Adam Klivans | **Alexandros G. Dimakis** | **Kristen Grauman** | **Jonathan I. Tamir** | **Daniel J. Diaz** | **Karen Davidson**

Department of Computer Science, The University of Texas at Austin, Austin, Texas, USA

Correspondence
Adam Klivans, Department of Computer Science, The University of Texas at Austin, Austin, TX, USA.
Email: karen_davidson@utexas.edu

Funding information
National AI Research Institutes program

Abstract

The Institute for Foundations of Machine Learning (IFML) focuses on core foundational tools to power the next generation of machine learning models. Its research underpins the algorithms and data sets that make generative artificial intelligence (AI) more accurate and reliable. Headquartered at The University of Texas at Austin, IFML researchers collaborate across an ecosystem that spans University of Washington, Stanford, UCLA, Microsoft Research, the Santa Fe Institute, and Wichita State University. Over the past year, we have witnessed incredible breakthroughs in AI on topics that are at the heart of IFML's agenda, such as foundation models, LLMs, fine-tuning, and diffusion with game-changing applications influencing almost every area of science and technology. In this article, we seek to highlight seek to highlight the application of foundational machine learning research on key use-inspired topics:

- Fairness in Imaging with Deep Learning: designing the correct metrics and algorithms to make deep networks less biased.
- Deep proteins: using foundational machine learning techniques to advance protein engineering and launch a biomanufacturing revolution.
- Sounds and Space for Audio-Visual Learning: building agents capable of audio-visual navigation in complex 3D environments via new data augmentations.
- Improving Speed and Robustness of Magnetic Resonance Imaging: using deep learning algorithms to develop fast and robust MRI methods for clinical diagnostic imaging.

IFML is also responding to explosive industry demand for an AI-capable workforce. We have launched an accessible, affordable, and scalable new degree program—the MSAI—that looks to wholly reshape the AI/ML workforce pipeline.

INTRODUCTION

The Institute for Foundations of Machine Learning (IFML) was part of the first cohort of the National Science Foundation's artificial intelligence (AI) institutes. We are home to a dynamic and diverse team of researchers focusing on machine learning foundations—the apparatus that underpins AI. Headquartered at The University of Texas at Austin, IFML researchers collaborate across an ecosystem that spans University of Washington, Microsoft Research, and Wichita State University. Together, we conduct foundational research that has the potential to impact every person on the planet. Over the past year, we have witnessed incredible breakthroughs in AI on topics that are at the heart of IFML's agenda: foundation models, generative models, fine-tuning, diffusion, algorithms, data augmentation, robustness, and reinforcement learning, to name a few. It was a thrilling year for our team, as we saw game-changing applications influencing almost every area of science and technology.

IFML works to understand the key foundational questions that need to be solved so that machine learning can continue its upward trajectory. Part of IFML's mission, as mandated by the NSF, is to function as a nexus point for other institutes and centers that may have overlapping interests in machine learning. We believe in the power of foundational research and its importance to both short and long-term innovation in machine learning. Within the current empirical framework, reducing the amount of trial and error using principled heuristics is extremely impactful. New algorithms and analyses have the potential to dramatically reshape the field. For example, the invention of polynomial-time interior-point methods for solving linear programs has influenced nearly every aspect of optimization, an area at the core of modern machine learning systems. Additionally, foundational researchers are now routinely hired by the largest technology companies. Given this tight integration, theoretical research is positioned to have major influence.

To meet future demand fora highly skilled AI workforce, IFML members were instrumental in developing coursework for a new Master of Science in Artificial Intelligence (MSAI) degree program at The University of Texas at Austin. The MSAI was featured in *The New York Times* due its potential to reshape the landscape of AI education.

The MSAI is explicitly designed to deliver affordability, accessibility, and scalability. These are the same traits that make this a uniquely valuable tool for workforce development:

- *Accessibility*: Because the program is online and part-time for most students, it can be completed without the student having to leave the workforce, relocating, or even making night and/or weekend trips to campus as is required by many executive education programs.
- *Affordability*: The program is intentionally priced to mitigate one of the most significant barriers to graduate study: prohibitively high tuition. Whereas traditional specialty MS programs are typically priced between \$500,000 and \$100,000, the MSAI program will be priced at approximately \$10,000.
- *Scalability*: In steady state, we expect to enroll more than 2500 MSAI students while graduating more than 700 students per year. We already have clear evidence of the viability of this delivery model through our online MS programs in Computer Science (CS) and Data Science (DS), which together currently enroll more than 3000 students. We began accepting applications for the MSAI on June 1 and received more than 2900 applications for the inaugural cohort.

RESEARCH HIGHLIGHT 1 – FAIRNESS IN IMAGING WITH DEEP LEARNING

We have all seen the moment in the movies—detectives gather around a computer screen staring at a blurry image captured from a store security camera. With just a few clicks, magical "*zoom and enhance*" software unblurs the fuzzy image, revealing the culprit's license plate or the face of the villain. It is still science fiction, but the reality is that Deep Learning methods are getting closer to realizing this dream. Modern AI techniques can render high-quality images from blurry and low-resolution samples. These techniques can transform computational photography, medical imaging, increase resolution, and accelerate MRI and microscopy.

In 2020, a deep learning generative model with ground breaking performance was posted on the web. The model could turn low-resolution images to high-quality photos. A user uploaded a low-resolution image of President Obama, to obtain an image that is now called "*White Obama*." This was a startling display of bias in AI imaging algorithms. The deep learning model was reconstructing images with predominantly white features and a heated debate arose across Twitter and other social media. Alex Dimakis, IFML co-director and a Professor in the Chandra Family Department of Electrical and Computer Engineering, led a team of researchers to investigate. *Was the source of the problem the bias in the training data, or in the reconstruction algorithm, and how could we fix it?*

The algorithm that generated these images, PULSE, is using a deep generative model called StyleGAN that can produce artificial faces (like the ones shown on

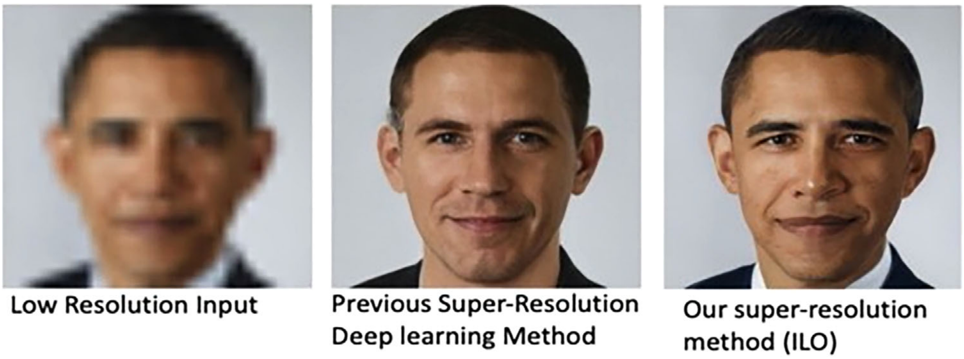

FIGURE 1 Low-resolution image input with comparisons of previous super-resolution deep learning method and the research team's super-resolution method.

thispersondoesnotexist.com) but is optimizing the generated image to match the low-resolution input image after downscaling. StyleGAN was trained on a dataset of images of predominantly white people. Many researchers argued that this was the source of the bias, and that simply creating a balanced training dataset would solve the problem. The team's recent work (Figure 1) shows that this is not the case: the reconstruction algorithm needs to be modified, in addition to the training set.

The first problem is that we needed to think carefully about defining fairness in reconstructing images of people with various attributes. Traditional group fairness definitions are defined with respect to specified protected groups—camouflaging the fact that these groupings are artificial and carry historical and political motivations. For instance, should South and East Asians be viewed as a single group or separate groups? Should we consider one race as a whole or further split by gender? Choosing which groups are valid and who belongs in them is an impossible dilemma and being "fair" with respect to Asians may require being "unfair" with respect to South Asians. This motivates our introduction of oblivious fairness. The machine learning algorithm needs to work for all possible groupings of the population.

Our first result is that several intuitive notions of group fairness are incompatible and impossible to achieve simultaneously. We show that the natural extension of demographic parity is strongly dependent on the grouping, and impossible to achieve obliviously. We introduce a new definition of fairness called Conditional Proportional Representation which can be achieved obliviously (i.e., without defining specific protected groups) through a natural algorithm that we propose.

Our second result is that the reconstruction algorithm previously used (MAP inference) is amplifying any bias that can be present in the data. The essence of why is illustrated in the following toy example: Alice flips a biased coin that comes Heads with probability 0.6 and Tails with 0.4. Bob has to guess Heads or Tails, knowing this is a biased coin. If Bob wants to maximize his probability of winning, he will always guess Heads, even when the bias is only 60%. This increases the 60% dataset bias to 100%, and we observe this bias amplification phenomenon experimentally in our study. On the contrary, if Bob uses the algorithm we propose (posterior sampling), Bob will randomize his guess and propose Heads only 60% of the time, matching but not amplifying the bias in the training data (Figure 2).

As deep learning imaging algorithms become ubiquitous across smartphones, social networks, MRI scanners, and a broad array of other applications, designing the correct metrics and understanding the foundations of deep learning is going to be key to ensure future deployments reflect the diverse and inclusive reality of our world.

RESEARCH HIGHLIGHT 2 – DEEP PROTEINS

One of the holy grails of biotechnology is the ability to engineer proteins by introducing mutations into their sequences in order to improve their function. To create protein-based biotechnologies, a thermostable scaffold is often needed for downstream industrial and therapeutic applications. For example, we may wish to thermostabilize the COVID-19 spike protein to improve the ability of our innate adaptive immune system to produce neutralizing antibodies. The computational stability prediction community, however, has struggled to find models that can generalize due to the large and highly complex nature of atomic-level data, as well as a lack of properly curated training sets. IFML members are focused on the fundamental problem of developing machine learning models that can predict stabilizing mutations on a given protein. Over the past year, IFML has developed new foundational methods for data augmentation, new representations and

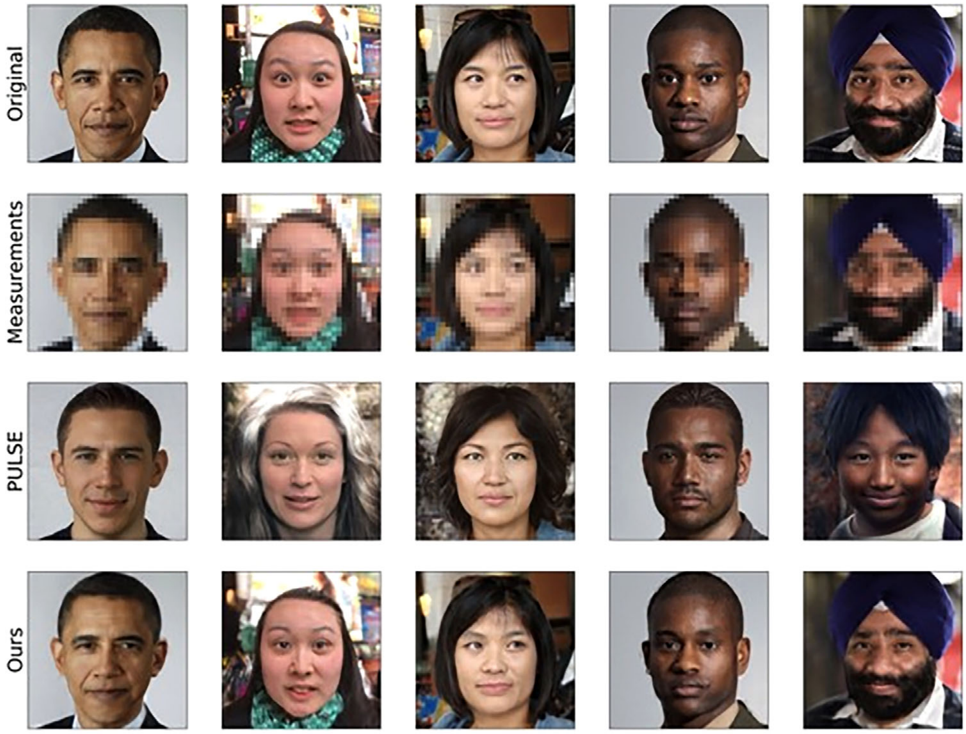

FIGURE 2 An example of reconstructions using posterior sampling compared to the previous method PULSE which uses MAP inference and amplifies the bias (Jalal et al., 2021a).

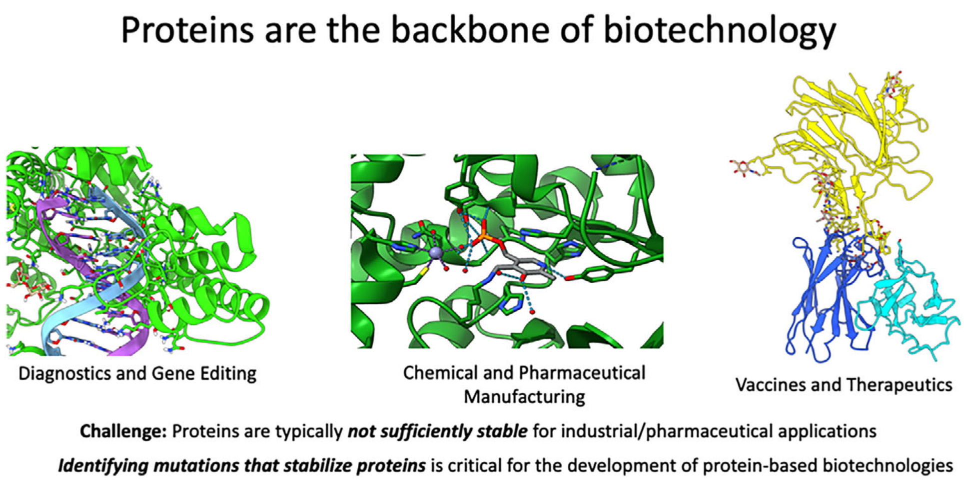

FIGURE 3 Identifying mutations that stabilize proteins can help advance innovations for industrial or pharmaceutical use.

algorithms for fast inference, and novel deep learning architectures (Figure 3).

The standard approach for building a thermostability predictor is to use a billion-parameter sequence model as a black box and fine-tune on a small-labeled training set. As our aim is to move beyond the limitations of these methods, we created new training sets using a simple and powerful data augmentation technique and leveraged both structure and sequence-based models to address both geometric and physical properties of proteins. The results required us to dig into foundational deep learning problems involving new architectures and representations. Our data augmentation technique opens the door for an order-of-magnitude increase in available data across a wide variety of protein engineering domains, as noted in "Stability Oracle: A Structure-Based Graph-Transformer for Identifying Stabilizing Mutations."

Due to experimental limitations, most datasets of mutations are biased toward particular amino acids. This makes it extremely difficult to make predictions on the entire space of mutations. Thermodynamic permutations (TP) is the first thermodynamic stability training set that sampled all 380 mutation types. We are currently extending this to higher order mutations. To address data leakage, we first generated a more comprehensive test set that better sampled the 380 mutation types and then used protein similarity to ensure no overlap between training test splits.

With these data enhancements in hand, we developed Stability Oracle, a structure-based stability predictor that, given a protein microenvironment and "to" and "from" amino acids, can rapidly predict the associated thermodynamic change. Prior structure-based work in thermostability prediction requires a complete structure of both the wildtype and mutated protein. This is computationally prohibitive, as it requires running AlphaFold to compute a single prediction. Further it has been empirically shown that AlphaFold struggles to update a protein structure based on a single mutation. Instead, we use a self-supervised pretraining phase based on previous work by Daniel Diaz on the MutCompute model but with an updated graph transformer architecture. This allows us to use representations of the "from" and "to" amino acids along with a microenvironment instead of obtaining entirely new structures. Fine tuning this Stability Oracle model using our TP data augmentation results in state-of-the-art thermostability prediction across all standard benchmarks and outperforms on all metrics.

Next, we discovered a way to extract "from" and "to" amino acid representations using AlphaFold, which is typically used for protein structure prediction. In fact, several publications have highlighted the inability of AlphaFold to capture changes in free energy of point mutations in proteins. Our approach and resulting model, StabilityFold, contradicts this conventional wisdom and demonstrates that Alphafold can be used to learn changes in free energy of point mutations. Further, we developed a new parallel algorithm called MutateEverything, which not only finetunes AlphaFold embeddings for single point mutations but can be generalized to high order mutations and can make millions of predictions using a single forward pass through AlphaFold's evoformer. Our method is quite general, and we can "plug-in" sequence-based models such as Prostata to obtain state of the art sequenced-based thermostability predictors that outperform for example ESM.

IFML is collaborating with UT Austin's Ellington lab to stabilize a lipase that can accelerate field diagnostics for Tuberculosis and stabilize a polymerase for synthetic biology applications. With the Georgiou Lab, we are stabilizing two human enzymes with applications in breast and colon cancer, respectively. We are working with the McClellan lab to accelerate the stabilization of several viral spike proteins for vaccine design. Finally, we are working with the Adapt lab at Houston Methodist Research Institute where we are working on improving the expression of COVID-19 antibodies and exploring the ability of Stability Oracle to predict mutations that improve binding affinity to the Omicron spike protein.

Long term, our goal is to create is a suite of deep learning tools that accelerate the engineering of different protein phenotypes and provide the computational foundation for the biomanufacturing revolution of the chemical, pharmaceutical, agrochemical industries.

RESEARCH HIGHLIGHT 3 – SOUNDS AND SPACE FOR AUDIO-VISUAL LEARNING

Humans learn by interacting with the world. IFML researchers have adopted a similar strategy to teach virtual agents new skills. We use all our senses when we navigate the world, but today's embodied AI agents—like robots or virtual assistants—are typically restricted to using only visual perception of their environment. IFML member Kristen Grauman, a Professor in the Department of Computer Science at UT Austin and a Research Scientist in Facebook AI Research (FAIR), is working to fill this void by building agents capable of audio-visual navigation in complex, acoustically and visually realistic 3D environments.

SoundSpaces is a first-of-its-kind audio-visual platform for embodied AI. We are working to produce realistic audio rendering based on room geometry, materials, camera position, and sound location— so we can create smart agents that can better respond to real-world situations. In this demo, we see the platform responding to a real-life scenario, such as a fire alarm going off during a piano lesson.

Building on their SoundSpaces platform, the team has developed multimodal deep reinforcement learning approaches to train navigation policies end-to-end from a stream of egocentric audio-visual observations. These policies allow the agent to (1) discover elements of the geometry of the physical space indicated by the reverberating audio and (2) detect and navigate to sound-emitting targets. Additionally, the project introduces a dataset of audio renderings based on geometrical acoustic simulations for two sets of publicly available 3D environments (Matterport3D and Replica), and extends Habitat to support the new sensor, making it possible to insert arbitrary sound sources in an array of real-world scanned environments. The research shows that audio greatly benefits from embodied visual navigation in 3D spaces and lays the groundwork for new research in embodied AI with audio-visual perception.

The latest iteration of SoundSpaces delivers on-the-fly geometry-based audio rendering for 3D environments. Given a 3D mesh of any real-world environment, SoundSpaces can generate highly realistic acoustics for arbitrary sounds captured from arbitrary microphone locations. Together with existing 3D visual assets, it supports an array of audio-visual research tasks, such as audio-visual navigation, mapping, audio source localization and separation, and visual-acoustic matching.

Compared to existing resources, SoundSpaces 2.0 has the advantages of allowing continuous spatial sampling, generalization to novel environments, and configurable microphone and material properties. This is the first geometry-based acoustic simulation that offers high fidelity and realism while also being fast enough to use for embodied learning. SoundSpaces and SoundSpaces 2.0 are publicly available to facilitate wider research for perceptual systems that can both see and hear.

RESEARCH HIGHLIGHT 4 – IMPROVING SPEED AND ROBUSTNESS OF MAGNETIC RESONANCE IMAGING

“Stress inducing” is often a phrase associated with magnetic resonance imaging (MRI) and more than 12 million people undergo the procedure each year. Jon Tamir, assistant professor in the Chandra Family Department of Electrical and Computer Engineering at UT Austin, works with a team of researchers to develop fast and robust MRI methods for clinical diagnostic imaging. Working with colleagues at UT Austin’s Dell Medical School, Tamir and his team are developing machine learning methods to shorten the times MRI exams can take while likewise extracting more data from the process.

MRI is an exceptional imaging modality because it does not use harmful radiation, but scan time is a major barrier to wider clinical adoption. Making the scanner faster and more comfortable will go a long way to serving a larger population, especially in pediatric and neonatal settings where a sick patient cannot be expected to stay still for very long.

Tamir and his team are using foundational AI algorithms developed at IFML to combine deep learning with biomedical imaging in a principled manner. A significant body of prior work focuses on developing end-to-end deep learning methods that reconstruct images from MRI measurements. The use of end-to-end deep learning methods is hindered by their black-box nature, due to lack of interpretability, trust, and robustness guarantees. Such methods perform extremely well when evaluated in carefully controlled environments but are fragile when used in clinical environments with disparate scanner hardware, imaging protocols, and patient populations. The reason for this fragility is the explicit coupling of the measurement model and the statistical image prior during training (Jalal et al., 2021b). In essence, a prescribed measurement operator and a dataset must both be specified at train time. When the test-time conditions differ, then the reconstructions will suffer from artifacts due to generalization error. In practice, clinical MRI demands the flexibility to change measurement settings on a per-patient basis to accommodate the natural heterogeneity in patient populations.

To overcome these limitations, the team designed an algorithm that explicitly decouples the measurement model and the statistical image prior. They used established and principled statistical physical models to separately describe the likelihood function of the measurement system. Then, they trained a foundation generative model on MRI images, similar to how Dall-E and Stable Diffusion were trained on images from the web. They combined the measurement model and the generative model and posed the image reconstruction task as sampling from the posterior distribution. As a result, they were able to accelerate the MRI scan by factors of $4x$–$12x$ while maintaining high image quality (Levac, Jalal, & Tamir, 2023). A benefit of this approach is the ability estimate uncertainty in the reconstruction, potentially providing more nuanced information about the quality of the result to the clinician. In addition to estimating uncertainty in the reconstruction, Tamir and coworkers extended their framework to incorporate uncertainty in the measurement model itself. For example, patient motion during the scan leads to ambiguity in the acquired measurements. By modeling the motion as an unknown random variable, they were able to extend the reconstruction to sample from the joint posterior of both the image and the motion parameters. Motion is a serious issue in MRI and this approach lets us get diagnostic images even when the patient moves during the scan.

Experimental scan performed by Tamir’s team at Dell Medical School on a healthy volunteer with institutional review board approval. The left image shows a brain MRI scan performed when the subject was still. The middle image is the result when the subject moved during the scan and shows artifacts near the ventricles. The right image is the proposed algorithm that removes the motion-induced artifacts (Figure 4).

SUMMARY

IFML research in use-inspired areas—imaging, video, navigation, and protein design/engineering—are the applications that have been at the heart of some exciting recent breakthroughs. IFML senior personnel are closely

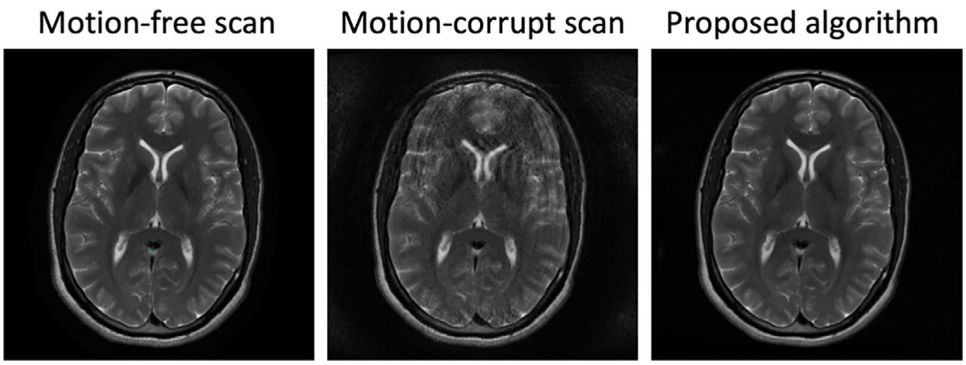

FIGURE 4 The research team extended their framework to incorporate uncertainty, such as patient movement, in the measurement model itself (Levac, Jalal, & Tamir, 2023).

connected to industrial partners in these areas, resulting in research that is aligned with the latest developments and has the potential for near-term deployment.

Our use-inspired work has become considerably more collaborative, as common foundational techniques (e.g., transformers) find applications across multiple modalities. As such, foundational research is now vital for achieving impact in diverse application areas. We use new algorithms, data sets, and architectures to ensure fairness in generative imaging, to enable robots to navigate in ever-changing environments, to deploy more robust MRI in clinical health settings, and to develop new biologics and therapeutics.

Leveraging our strong partnerships in education and broadening outreach efforts, IFML continues to address the demand for an increasingly AI-centric workforce. We strive to be a leader in AI education by providing a globally available, low-cost online Master of Science in AI.

ACKNOWLEDGMENTS

This material is based upon work supported by the National AI Research Institutes program supported by NSF under Award No. 2019844. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict.

ORCID

Karen Davidson https://orcid.org/0009-0000-4528-7998

REFERENCES

Jalal A., S. Karmalkar, J. Hoffmann, A. G. Dimakis, and E. Price. 2021a. "Fairness for Image Generation with Uncertain Sensitive Attributes." In International Conference on Machine Learning (ICML).

Jalal A., M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir. 2021b. "Robust Compressed Sensing MRI with Deep Generative Priors." In Advances in Neural Information Processing Systems (NeurIPS).

Levac B., A. Jalal, and J. I. Tamir. 2023. "Accelerated Motion Correction for MRI Using Score-Based Generative Models." In IEEE International Symposium on Biomedical Imaging (ISBI).

How to cite this article: Klivans, A., AG. Dimakis, K. Grauman, JI. Tamir, DJ. Diaz, and K. Davidson. 2024. "*Institute for Foundations of Machine Learning (IFML): Advancing AI systems that will transform our world.*" *AI Magazine* 45: 35–41. https://doi.org/10.1002/aaai.12163

AUTHOR BIOGRAPHIES

Adam Klivans is IFML director and a professor in the Department of Computer Science at UT Austin.

Alexandros G. Dimakis is IFML co-director and a professor in the Chandra Family Department of Electrical and Computer Engineering at UT Austin.

Kristen Grauman is a professor in the Department of Computer Science at UT Austin.

Jonathan I. Tamir is an assistant professor in the Chandra Family Department of Electrical and Computer Engineering at UT Austin.

Daniel J. Diaz is an IFML postdoctoral researcher at UT Austin.

Karen Davidson is communications coordinator for IFML and the Machine Learning Lab at UT Austin.