Potential Synergies in Physics, AI, and Medical Imaging

Mike Williams

Department of Physics
NSF AI Institute for Artificial Intelligence and Fundamental Interactions
Laboratory for Nuclear Science
Statistics & Data Science Center
MIT

March 17, 2023

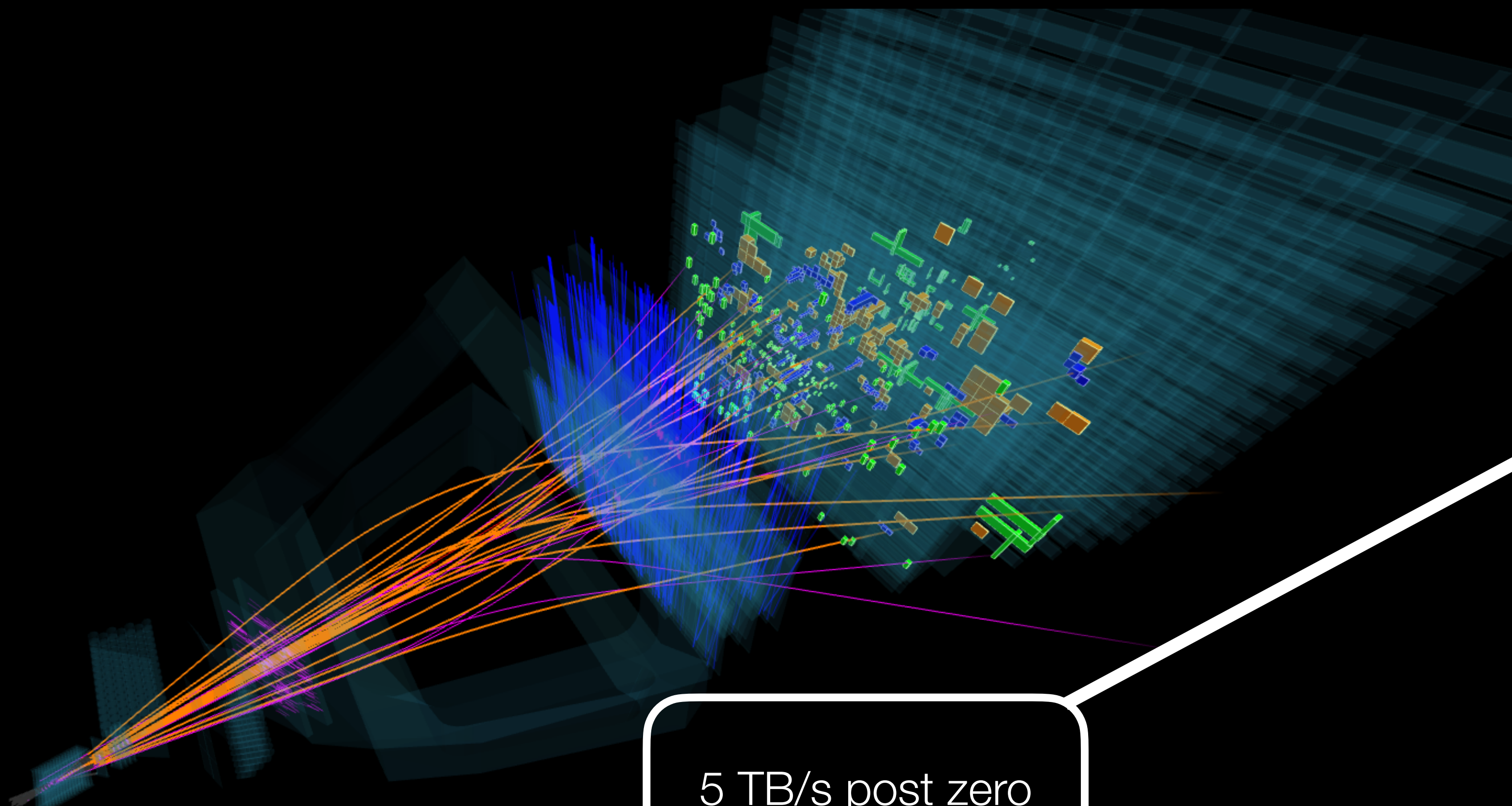
Institute for Artificial Intelligence and Fundamental Interactions (*IAIFI /ai fai/ https://iaifi.org*)

Enable physics discoveries by developing and deploying the next generation of AI technologies

Galvanize AI research innovation by incorporating physics intelligence into artificial intelligence

NSF | MIT | HARVARD | NORTHEASTERN UNIVERSITY | TUFTS

Making Decisions @ 40 MHz
(and living with the consequences)



5 TB/s post zero suppression (30 EB / year)

All collisions processed in real time on GPUs to infer what particles were produced and what their properties were. Mixture of traditional and AI algorithms used.

Vast majority of data must be discarded. AI used to make most of these decisions.

Data analyzed later by physicists. Mixture of AI and traditional methods used to produce published results.

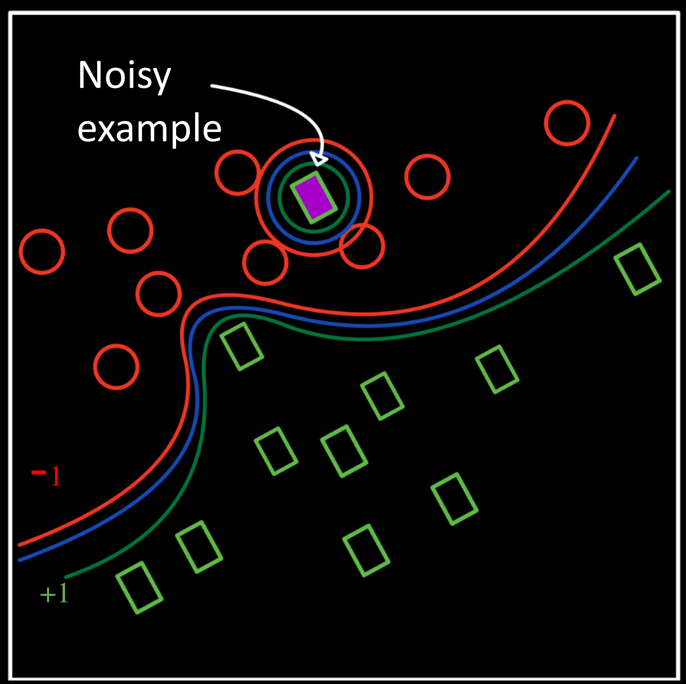
Algorithms used in the real-time environment must be **robust** and **interpretable** — they must account for detector resolution, instability, known unknowns, and must provide formal behavioral/performance **guarantees** to convince us that they are fit for purpose.

In short, we must be able to **trust** them to make important irreversible decisions.

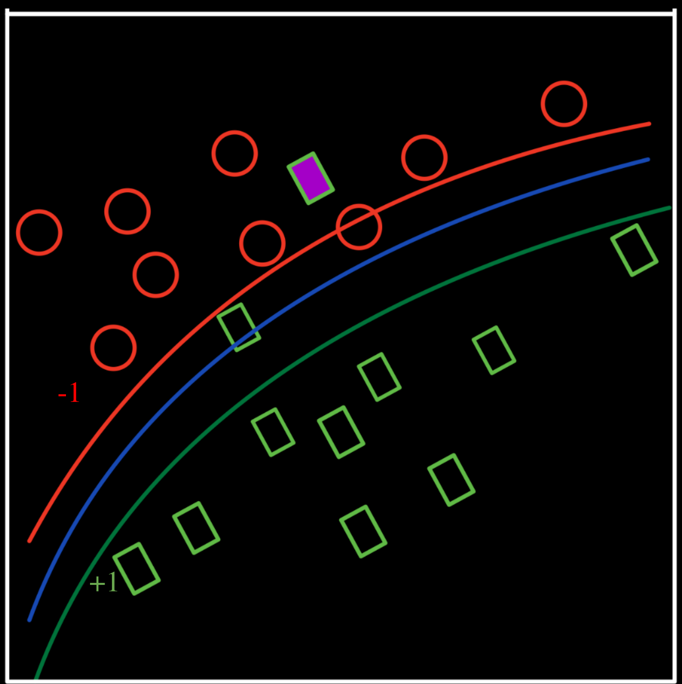
Robust AI

Neural networks can be universal function approximators even in high dimensions, which allows them to solve some incredibly hard problems — but in the real world our ideal solution is NOT found in the set of all functions, but a restricted set of robust ones.

Deep NN overfits on training noise



Robust NN respects resolution scale, etc.



Our solution was to create an architecture that guarantees a **bound on the gradient** of the learned function in each direction in feature space:

Domain expert specifies a priori **inductive bias** on learning scales.

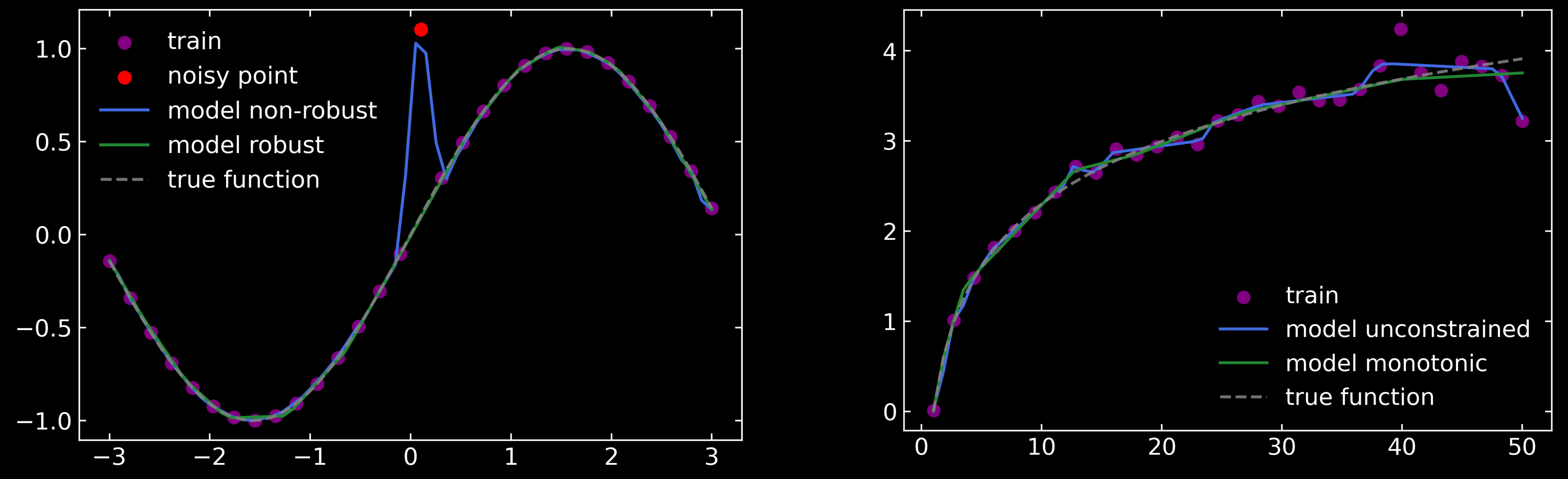
$$\left|\frac{\partial F}{\partial x_i}\right| \leq \lambda \rightarrow |F(\vec{x} + \vec{\epsilon}) - F(\vec{x})| \leq \lambda \|\vec{\epsilon}\|_1$$

Furthermore, we can also make the learned function **monotonic** in any feature direction by simply adding a linear function in that direction!

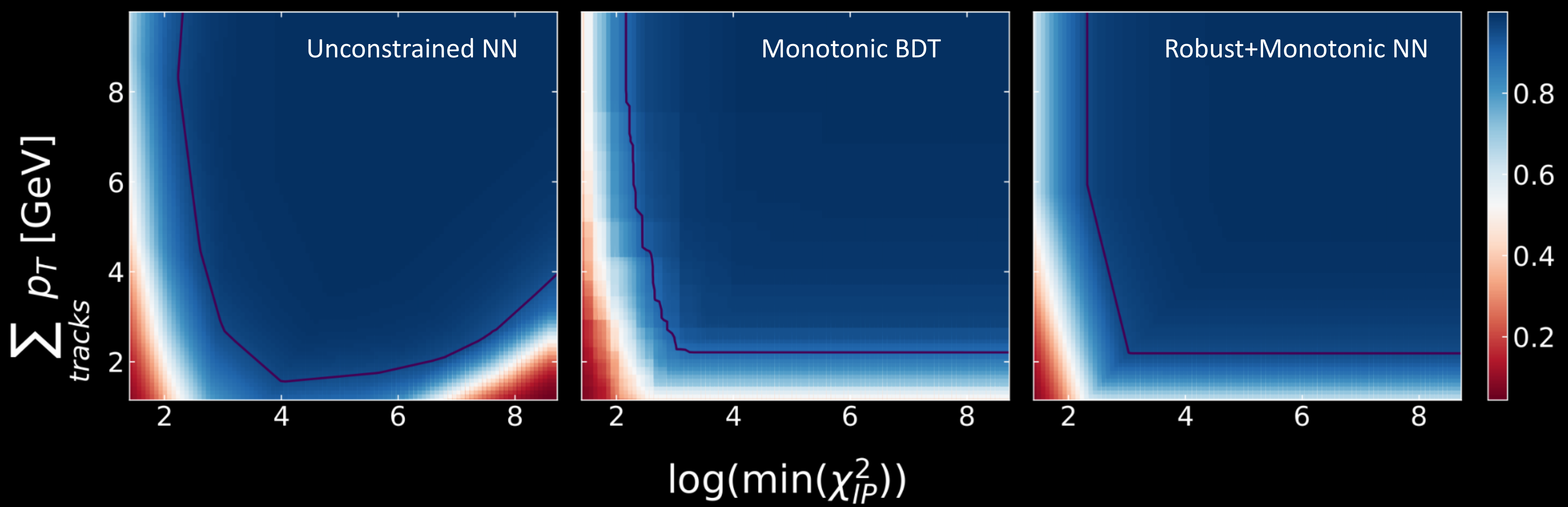
Kitouni, Nolte, MW [NeurIPS 2021, 2112.00038]

Robust AI

Toy demonstrations that our algorithm works as expected.



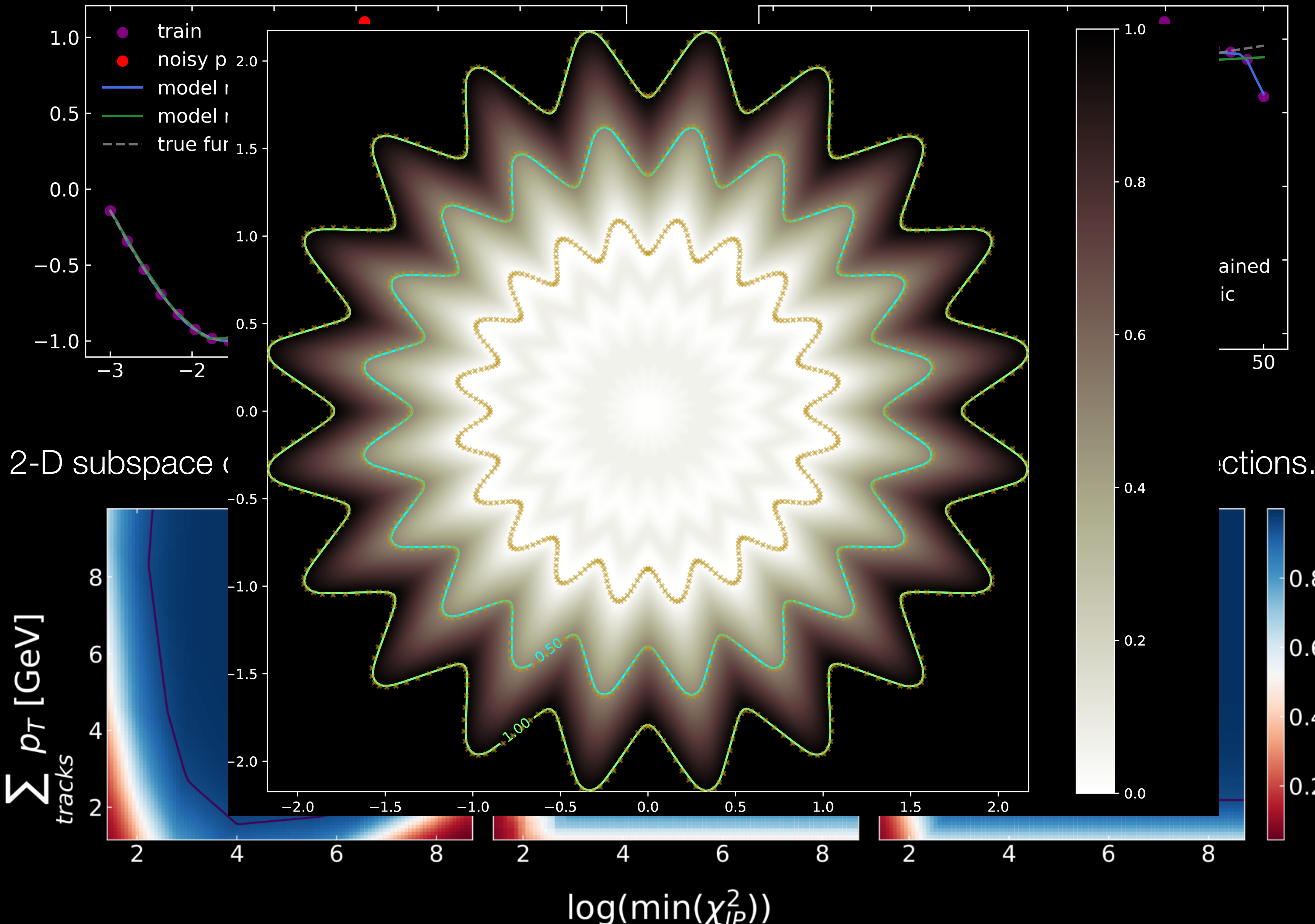
2-D subspace of the LHCb trigger selection. Our algorithm adopted for all major selections.



Kitouni, Nolte, MW [NeurIPS 2021, 2112.00038]

Robust AI

Toy demonstrations that our algorithm works as expected.



Kitouni, Nolte, MW [NeurIPS 2021, 2112.00038]

Robust & Monotonic AI Applications

We applied our LHC technology out of the box to various benchmark problems where some features are known to be monotonic, and we beat state-of-the-art models everywhere — with tiny networks!

Kitouni, Nolte, MW [ICLR 2023]

COMPAS

Method	Parameters	$\upuparrows$ Test Acc
Certified	23112	$(68.8 \pm 0.2)\%$
LMN	**37**	$(\mathbf{69.3 \pm 0.1})\%$

BlogFeedback

Method	Parameters	$\downdownarrows$ RMSE
Certified	8492	$.158 \pm .001$
LMN	**2225**	$\mathbf{.160 \pm .001}$
LMN mini	**177**	$\mathbf{.155 \pm .001}$

LoanDefaulter

Method	Parameters	$\upuparrows$ Test Acc
Certified	8502	$(65.2 \pm 0.1)\%$
LMN	**753**	$(\mathbf{65.44 \pm 0.03})\%$
LMN mini	**69**	$(\mathbf{65.28 \pm 0.01})\%$

ChestXRay

Method	Parameters	$\upuparrows$ Test Acc
Certified	12792	$(62.3 \pm 0.2)\%$
Certified E-E	12792	$(66.3 \pm 1.0)\%$
LMN	**1043**	$(\mathbf{67.6 \pm 0.6})\%$
LMN E-E	**1043**	$(\mathbf{70.0 \pm 1.4})\%$

Heart Disease

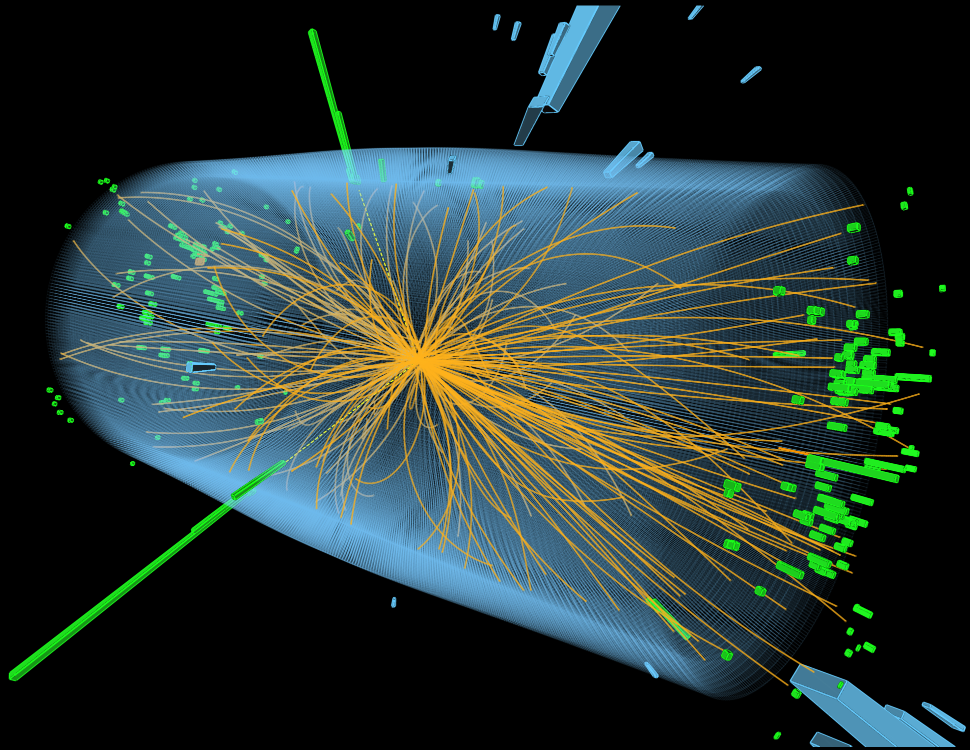
Method	$\upuparrows$ Test Acc
COMET	$(86 \pm 3)\%$
LMN	$(\mathbf{89.6 \pm 1.9})\%$

Auto MPG

Method	$\downdownarrows$ MSE
COMET	$(8.81 \pm 1.81)\%$
LMN	$(\mathbf{7.58 \pm 1.2})\%$

ChestXRay diagnoses diseases visible in a chest x-ray, required to be monotonic in age and number of patient follow ups. Heart Disease predictions are monotonic in blood pressure and cholesterol.

FastML



150 TB/s post zero suppression (1 ZB / year)

Vast majority of data must be discarded within a few microseconds. AI developed to make decisions in under 100 nanoseconds!

Selected collisions processed in real time on (C,G)PUs to infer what particles were produced and what their properties were. Mixture of traditional and AI algorithms used.

Data analyzed later by physicists. Mixture of AI and traditional methods used to produce published results.

For CMS and ATLAS, the data volumes are too large to read out at 40MHz. Algorithms must decide which events to keep before the data buffers fill up, about 4 microseconds.

Extremely **low-latency inference** is essential! Many novel tools developed by LHC collaborators (founding member Phil Harris is in IAIFI and co-leads A3D3), especially for running **AI on FPGAs**.

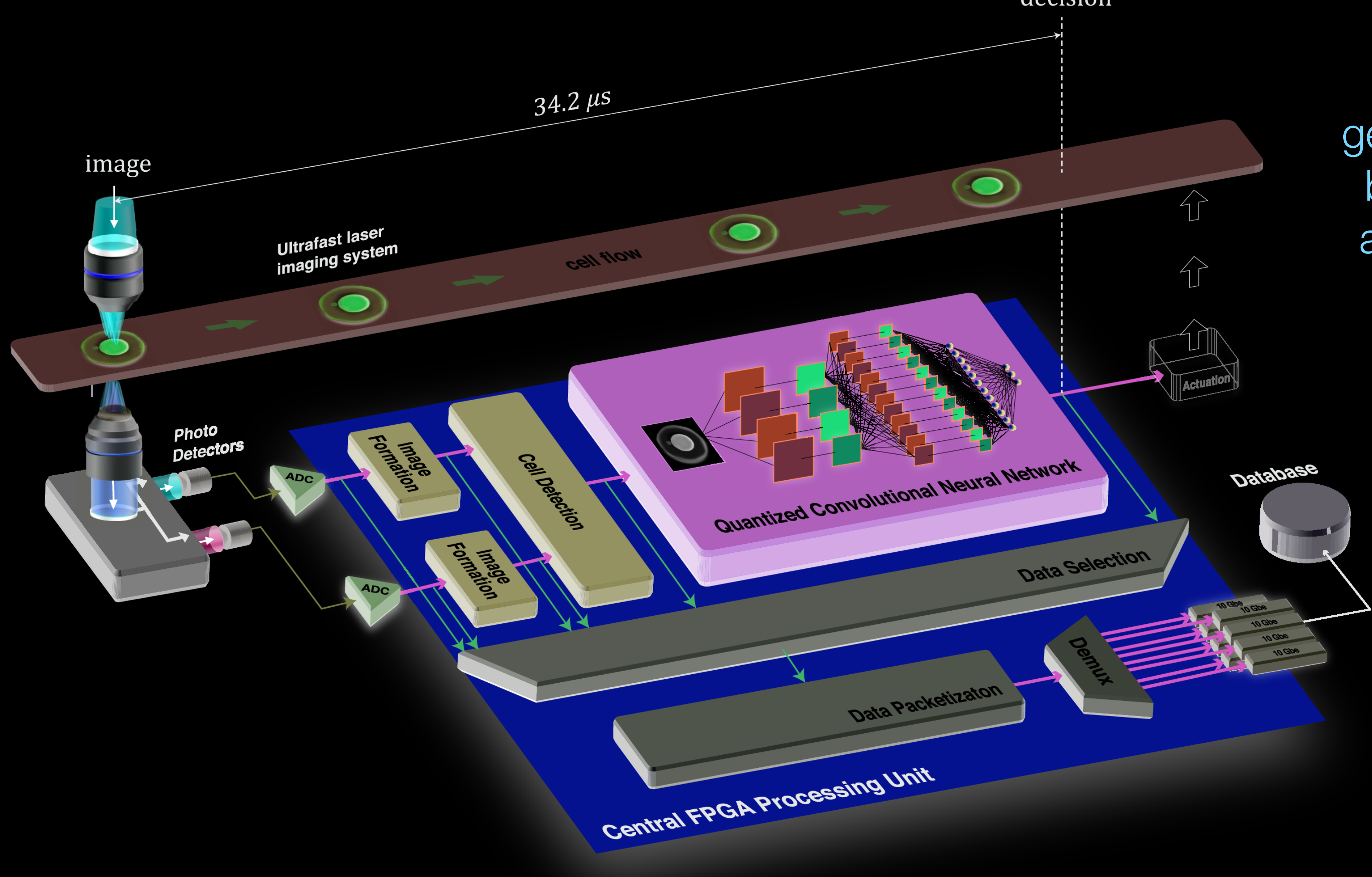
https://fastmachinelearning.org

FastML

Potential FastML applications in medicine include anywhere where super low-latency is required.

[Wang + [IEEE 2022]]

FastML developed for the LHC is completely generic and could be used almost anywhere out of the box.



Example of a reconfigurable FPGA for intelligent real-time image analysis which performs image cell detection and classification using AI with the ultra low latency needed to tag the cells.

A3D3

Accelerated AI Algorithms for Data-Driven Discovery funded by the NSF HDR program to expand on FastML to apply real-time AI at scale to advance scientific knowledge and accelerate discovery.

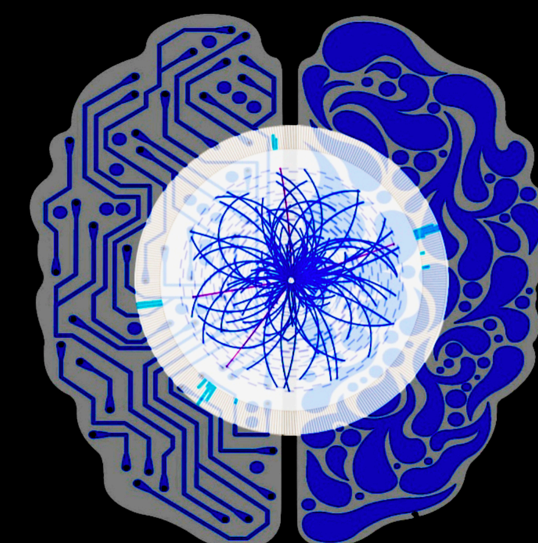
https://a3d3.ai

Hardware + Algorithm Codevelopment

Domain scientists, AI, and hardware experts creating and applying algorithms that exploit structured scientific data and hardware-AI co-design.

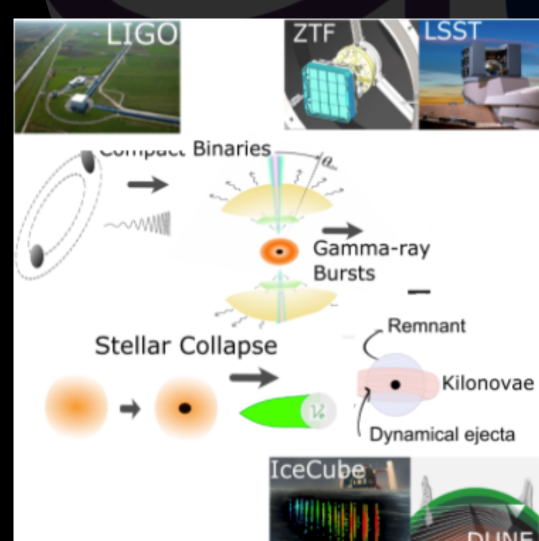
High Energy Physics

Working to deploy fully functional heterogeneous AI accelerated High Level Trigger systems for the LHC experiments.



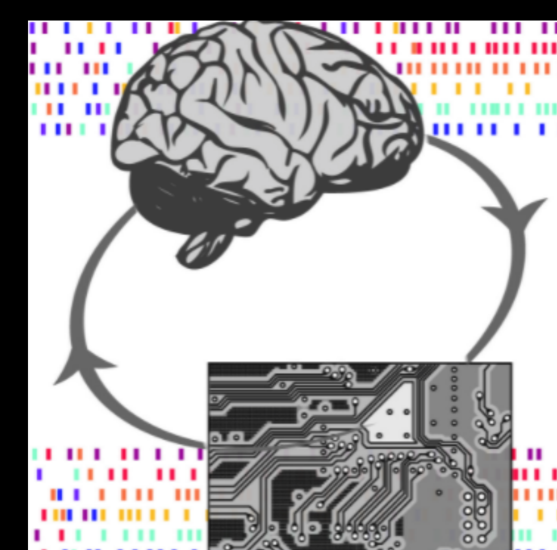
Multi-Messenger Astronomy

R&D on low-latency infrastructures to enable the study of objects previously inaccessible to AI-based studies.



Neuroscience

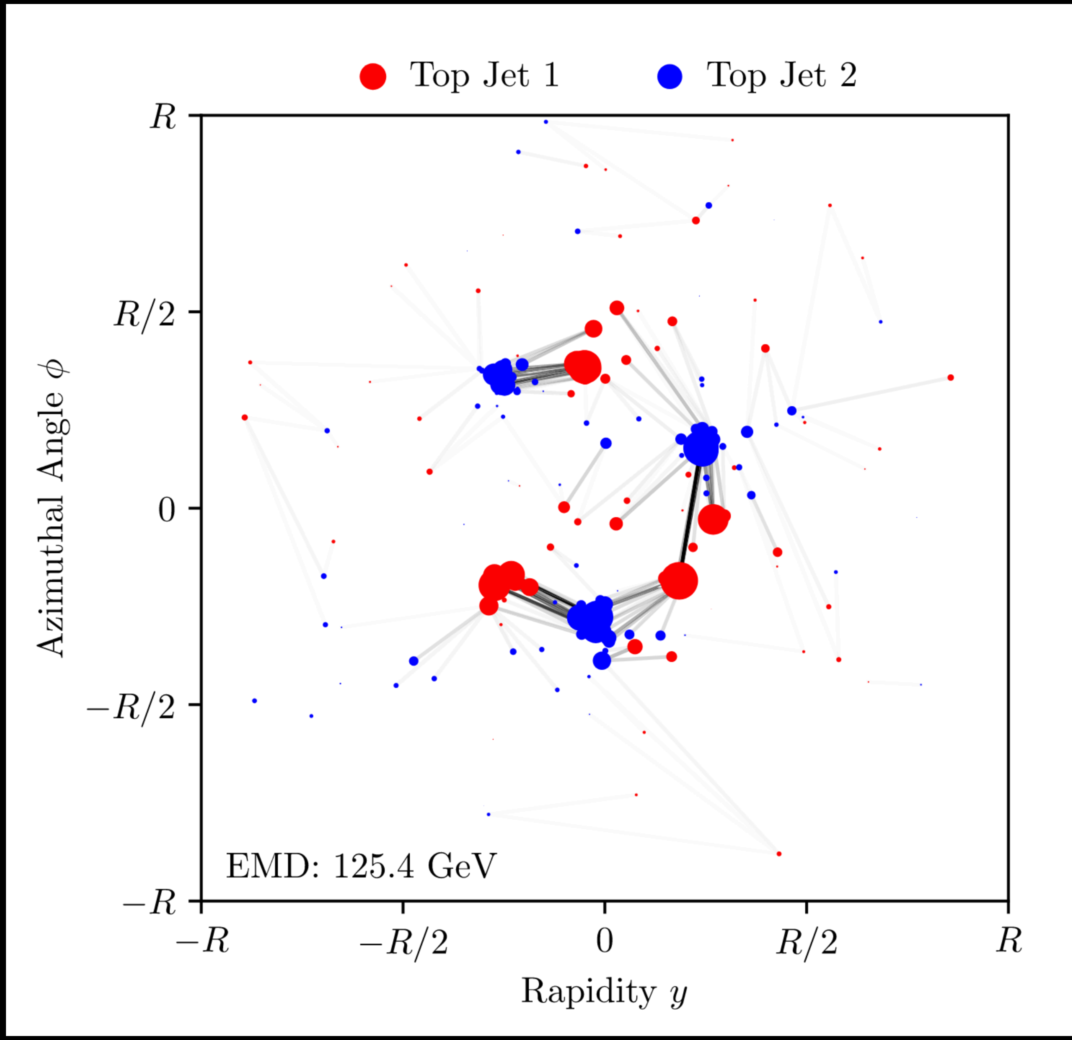
Developing AI methods that are high-throughput and low-latency to enable significant advances in our understanding of brain function.



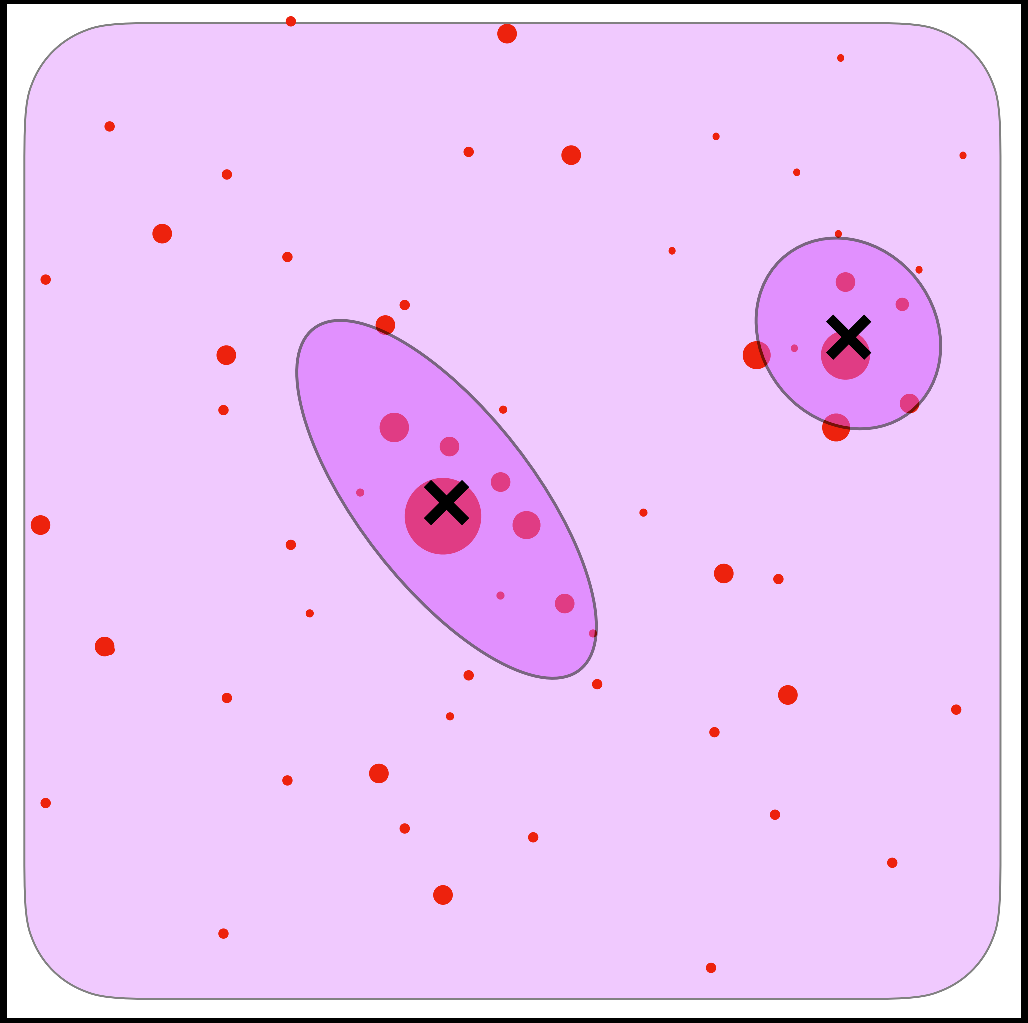
Earth Mover's Distance

The Earth Mover's Distance (**Wasserstein metric**), a key component of Optimal Transport theory, determines a distance between 2 distributions. Recently, it was shown that the EMD (probability replaced by energy) is the natural metric for the space of particle-collider events.

Komiske, Metodiev, Thaler [PRL, 1902.02346]



Ba, Dogra, Gambhir, Tassisa, Thaler [NeurIPS 2022, 2302.12266]



Now it is possible to do regression between point-cloud data and parametrized manifolds using the EMD as loss. Could this same idea be applied to medical imaging to not only detect the presence of unhealthy tissue but to determine its properties?

See also Kitouni, Nolte, MW [NeurIPS 2022, 2209.15624]

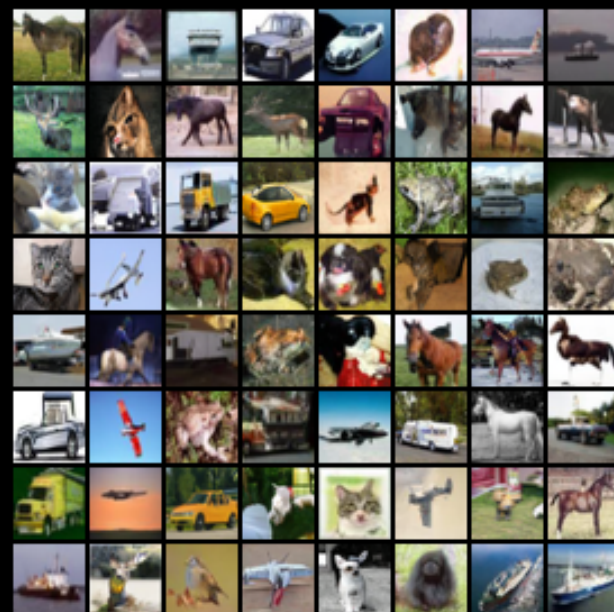
Et Cetera

Even just within the IAIFI there are too many interesting and relevant projects to mention in this talk.

Physics-Inspired Generative Models

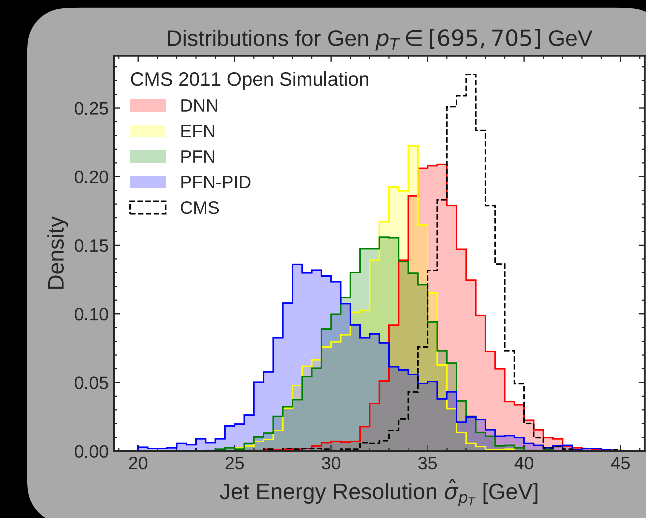
PFGM++ unifies diffusion models (thermodynamics) and Poisson flows (electrodynamics).

Xu, Liu, Tian, Tong, Tegmark, Jaakola [2302.04265]



Machine Learning Uncertainties

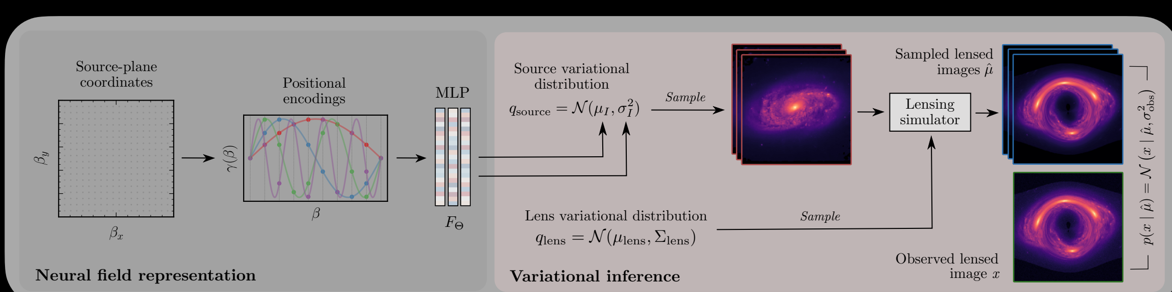
Calibration by inferring corrections, uncertainties, and correlations simultaneously using ML leads to improved measurements.



Gambhir, Nachman, Thaler [PRL, 2205.03413]

Strong Lensing Source Reconstruction

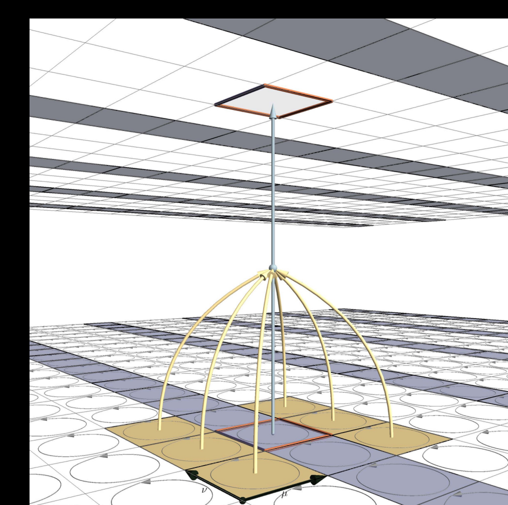
Probability distributions of source images obtained non-parametrically.



Mishra-Sharma, Yang [ICML, 2206.14820]

Exact Symmetries in Normalizing Flows

Exact symmetry equivariant normalizing flows developed for advanced QCD calculations could easily be adapted for other applications (e.g. exact rotational equivariance).

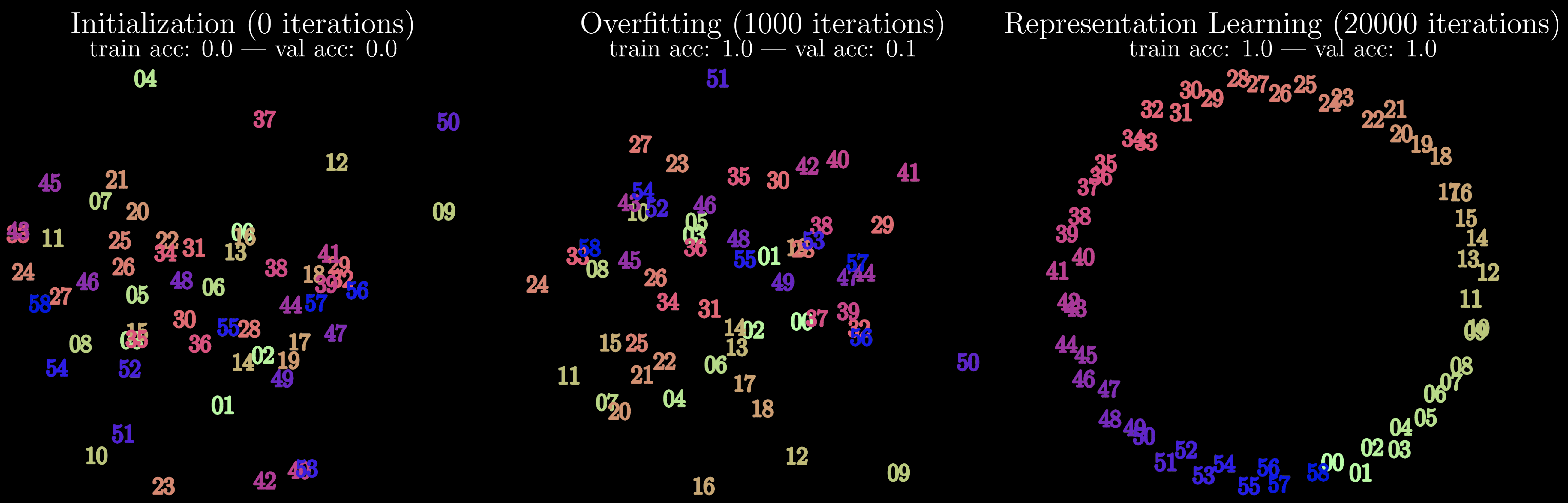


IAIFI, DeepMind, + [2208.03832]

Deep Learning + Deep Thinking ➜ Deep Understanding

In simple problems, e.g. algorithmic ones, AI may seem to do very strange things, but in some cases we have been able to use physical reasoning to not only gain qualitative understanding, but to also develop theories that can quantitatively describe the learning dynamics.

Liu, Kitouni, Nolte, Michaud, Tegmark, MW [Oral Highlight @ NeurIPS 2022, 2205.10343]



We are starting to explore more directly the connections between a deeper understanding of how AI methods really work — and the field of AI Safety. We're open to collaborations with NIH!

Summary

- Building inductive biases into AI/ML methods not only leads to behavioral guarantees, but also typically to improved performance and less data needed in training.
- LHC running conditions have forced us to develop robust and interpretable AI — along with ultra low-latency / high-throughput AI.
- These technologies are already being applied to other domains, including health and neuroscience.
- Recent advances have made it possible to do parametric regression using the Earth Mover's Distance / Wasserstein metric, which is used in many image-processing tasks (and in optimal transport). Could this same idea be used in medical imaging?
- Many other IAIFI products likely have direct applications in health. (I showed a few.)
- Another major goal of the IAIFI is to combine deep thinking with deep learning to obtain a deeper understanding of how AI works — which we believe will ultimately greatly improve AI safety.

Please feel free to contact us at iaifi@mit.edu (or me directly at mwill@mit.edu) with any questions about our IAIFI research, Fellows program, Summer School, etc. We are also open to hosting a dedicated workshop at MIT on the intersections of Physics, AI, medical imaging, and AI Safety.

Thanks! Questions?

Are there synergies between physics, AI, and medical imaging?

There are several synergies between the fields of physics and AI, especially in the context of medical imaging. Here are a few examples:

1. Image acquisition and processing: Medical imaging technologies, such as CT, MRI, and PET, rely on complex physics principles to acquire and process images. Physics experts can collaborate with AI researchers to develop new imaging techniques and algorithms that can help improve the quality and resolution of medical images.
2. Algorithm development and optimization: Many AI algorithms used in medical imaging rely on advanced mathematical concepts and optimization techniques that have their roots in physics. Physicists can bring their expertise in these areas to help optimize and improve AI algorithms for medical imaging.
3. Machine learning and deep learning: Machine learning and deep learning, two of the key technologies underlying AI, rely heavily on statistical modeling and data analysis. These are areas where physicists also have significant expertise, making them valuable collaborators for AI researchers.
4. New imaging modalities: Physics researchers are constantly developing new imaging technologies, such as quantum imaging and terahertz imaging, which have the potential to revolutionize medical imaging. AI can be used to help analyze and interpret the large amounts of data generated by these new imaging modalities, enabling better diagnoses and treatment planning.

Overall, the synergies between physics and AI in the field of medical imaging hold great promise for advancing our understanding of the human body and improving patient care.

Are there concerns with using AI in medical imaging?

Yes, there are several concerns with using AI in medical imaging, including:

1. Data quality and bias: AI algorithms are only as good as the data they are trained on. If the data is incomplete or biased, the algorithm may make incorrect predictions or diagnoses. For example, if an AI algorithm is trained on a dataset that only includes images of white patients, it may not perform as well on images of patients of different races.
2. Interpretability and transparency: AI algorithms can be difficult to interpret, which can make it hard for doctors to understand how the algorithm arrived at a diagnosis. Lack of transparency can also make it difficult for regulators to evaluate the safety and effectiveness of AI algorithms.
3. Legal and ethical concerns: The use of AI in medical imaging raises legal and ethical concerns, particularly around data privacy and patient consent. For example, if an AI algorithm is used to diagnose a patient without their knowledge or consent, it could violate their privacy rights.
4. Integration with existing healthcare systems: Integrating AI into existing healthcare systems can be a complex and costly process, particularly if the system was not designed to support AI. This can lead to compatibility issues and delays in implementing new AI technologies.
5. Overreliance on AI: There is a risk that doctors may become over-reliant on AI and stop using their own clinical judgement. This could result in missed diagnoses or incorrect treatments. Therefore, it is important to ensure that AI is used as a tool to assist doctors rather than replace them.