Simulation AI:

State of the art, best practices, trends

Two Sentiments

YASAI

The Revolution

Omnipresence

Leaps in model ability

New models every month

The Mirage

Not much seems to work

Unclear what AI is supposed to do

Negative ROI

The Revolution

YASAI

AI is **All Around Us:**

AI Generated Image
Workload: 2 seconds typing

- Text / Image / Video Generation
- Programming Tools
- Conference Presentations

The Revolution

AI in Simulation:

The Revolution

AI in Simulation:

•Surrogate models for faster simulation results
•Near-real-time design evaluation
•Reduced-order models
•Neural operators for field prediction
•Geometry-to-field prediction
•Meshless flow/thermal prediction
•Fast pressure, temperature, stress, or velocity estimates
•Design-space exploration
•Optimization acceleration
•Inverse design
•Parameter studies
•Sensitivity analysis
•Uncertainty quantification
•Out-of-distribution detection
•Simulation result interpolation
•Simulation result extrapolation
•Digital twins
•Real-time monitoring models
•Predictive maintenance models
•Calibration of simulation models
•Model updating from test data
•Test/simulation correlation
•Material model identification
•Turbulence closure modeling
•Subgrid-scale modeling
•Boundary-condition estimation
•Initial-condition estimation
•Automatic mesh generation
•Geometry cleanup
•CAD-to-simulation preparation
•Automated boundary labeling
•Simulation setup assistants
•Parameter recommendation
•Failure detection in simulation runs
•Convergence prediction
•Early stopping of bad simulations
•Error estimation
•Result plausibility checks
•Anomaly detection in fields
•Feature extraction from simulation results
•Automated postprocessing
•Report generation
•Engineering decision support
•Requirements-to-simulation mapping
•Knowledge capture from past projects
•Search over simulation archives
•Similar-case retrieval
•Reuse of historical simulation data
•Multi-fidelity modeling
•Combining 1D, 3D, test, and field data
•Generative design
•Physics-informed machine learning
•Physics-constrained prediction
•Data-driven correction of low-fidelity models
•Hybrid simulation/ML workflows
•AI copilots for CAE tools
•Natural-language access to simulation data
•Automatic documentation of simulation workflows
•Risk screening before expensive simulations

AI Race

New papers present better models every month:

- Transolver++ (2 /2025)
- GFocal (8 / 2025)
- PGOT (12 /2025)
- MGN-T (1 /2026)
- Transolver 3 (2 /2026)

Model	Structured Mesh			Point Cloud
	Plasticity	Pipe	Airfoil	Elasticity
WMT (2021)	0.0076	0.0077	0.0075	0.0359
U-FNO (2022)	0.0039	0.0056	0.0269	0.0239
Geo-FNO (2023a)	0.0074	0.0067	0.0138	0.0229
U-NO (2023)	0.0034	0.0100	0.0078	0.0258
F-FNO (2023)	0.0047	0.0070	0.0078	0.0263
LSM (2023)	0.0025	0.0050	0.0059	0.0218
Galerkin (2021)	0.0120	0.0098	0.0118	0.0240
HT-Net (2024)	0.0333	0.0059	0.0065	/
OFormer (2023c)	0.0017	0.0168	0.0183	0.0183
GNOT (2023)	0.0336	0.0047	0.0076	0.0086
FactFormer (2023d)	0.0312	0.0060	0.0071	/
ONO (2024)	0.0048	0.0052	0.0061	0.0118
IPOT (2024)	0.0033	/	0.0088	0.0156
Transolver (2024)	0.0013	0.0043	0.0053	0.0079
SAOT (2025)	0.0014	0.0062	0.0051	0.0090
PGOT (Ours)	**0.0012**	**0.0039**	**0.0046**	**0.0069**
Improvement	7.7%	9.3%	9.8%	12.7%

Table 4. Relative L2 errors (in %) of surface pressure p_s and skin friction coefficient C_f on the NASA-CRM dataset, and surface pressure p_s, volume velocity u, wall shear stress τ and volume pressure p_v on the AhmedML and DrivAerML datasets.

Models	NASA-CRM		AhmedML				DrivAerML			
	p_s	C_f	p_s	u	τ	p_v	p_s	u	τ	p_v
Graph U-Net*	15.85	15.61	6.46	4.15	7.29	5.18	16.13	17.98	27.84	20.51
GINO*	12.39	11.51	7.90	6.23	8.18	8.80	13.03	40.58	21.71	44.90
GAOT*	30.38	59.79	8.02	7.43	9.92	10.47	34.00	57.18	61.00	56.90
UPT	12.78	23.78	4.25	2.73	5.80	3.10	7.44	8.74	12.93	10.05
AB-UPT	9.77	6.43	3.97	1.94	5.60	**2.07**	3.82	5.93	7.29	6.08
Transolver*	9.61	7.04	3.20	1.81	4.85	2.41	4.81	6.78	8.95	7.74
Transolver++*	9.51	6.95	3.47	1.78	5.06	2.35	4.12	4.70	6.42	6.70
Transolver-3	**8.71**	**5.85**	**2.96**	**1.60**	**4.81**	2.16	**3.71**	**4.14**	**5.85**	**5.72**

The Mirage

Why do we see so little true impact?

The Mirage

Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations

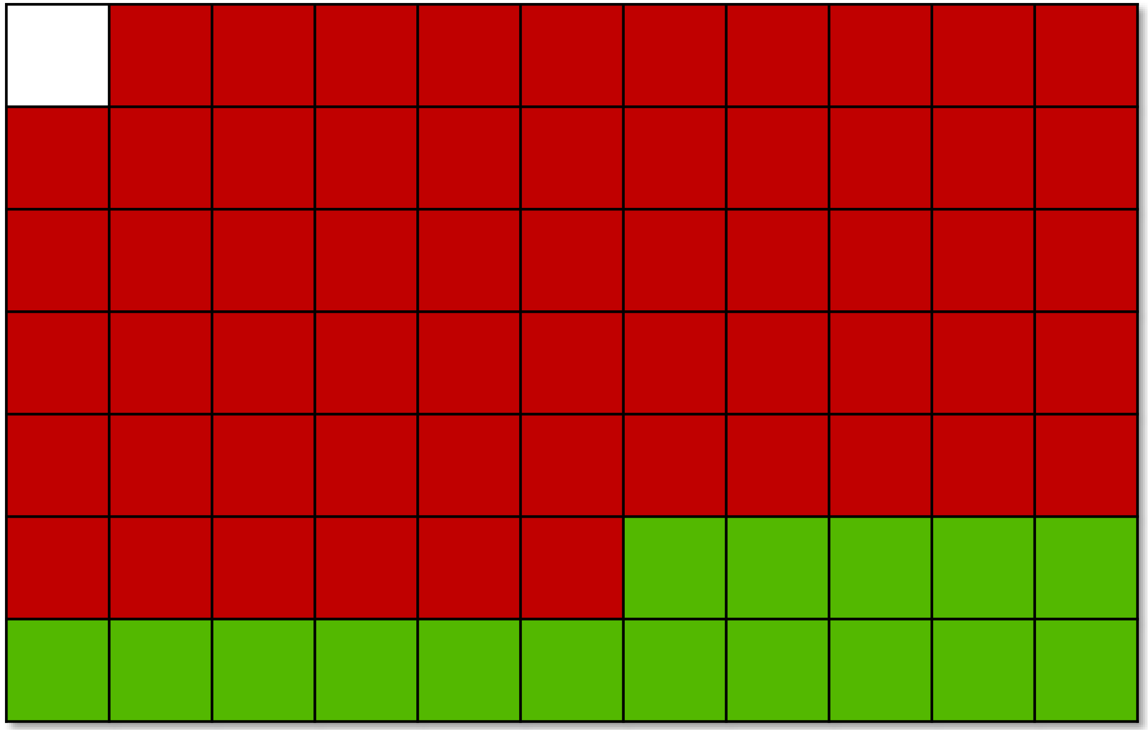
Nick McGreivy[1,2*] and Ammar Hakim[2]

[1*]Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey, USA.
[2]Princeton Plasma Physics Laboratory, 100 Stellarator Rd, Princeton, New Jersey, USA.

The Mirage

Finding:

60/76 papers publish **overoptimistic** results



What's real?

Hype

- 60 papers were overly optimistic or not reproducible

- An additional 6 papers published no improvement as result

Substance

- 16 papers did show **reproducible improvements** compared to standard numerical methods

Basic Model Types

Plain Neural Networks:

Multilayer Perceptrons

Convolutional Networks (CNN)

Recurrent Networks (RNN)

Graph Neural Networks:

GraphSAGE

GCN

GAT

Deep Learning

Transformers:

GPT

Transolver

Neural Operators:

DeepONet

FNO

Core Concepts

Convolution	Memory	Attention	Physics Information
• Spatial Correlations	• Sequential Correlations	• Fully learned set-wise correlations	• Additional equations

Application

Research and production-ready models are usually derived from:

Plain Neural Networks:

Multilayer Perceptrons

Convolutional Networks (CNN)

Recurrent Networks (RNN)

Graph Neural Networks:

GraphSAGE

GCN

GAT

Deep Learning

Transformers:

GPT

Transolver

Neural Operators:

DeepONet

FNO

+

Convolution
- Spatial Correlations

Memory
- Sequential Correlations

Attention
- Fully learned set-wise correlations

Physics Information
- Additional equations

Application

Most model architectures are a blend of basic algorithms and broader concepts

Example:

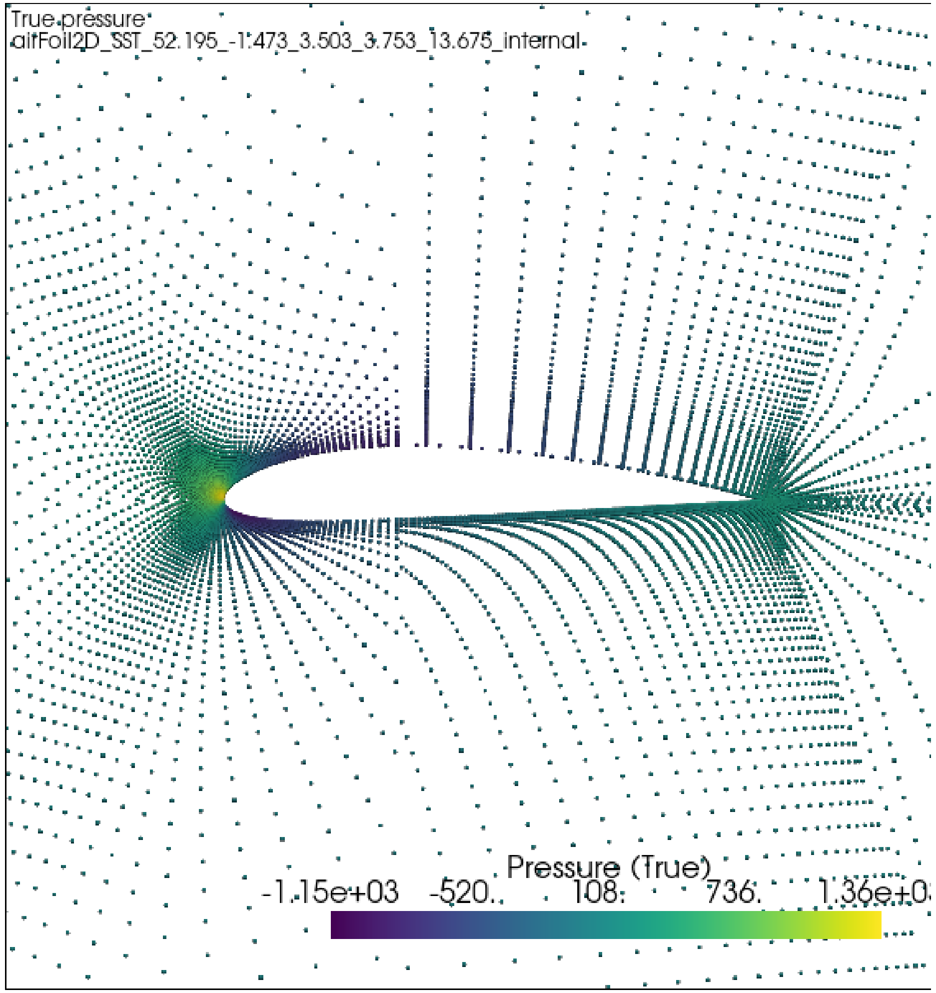
Transolver

Transolver

1. Groups datapoints into **slices with physical similarities**

2. An **attention function** weights the influence of a slice on all points

3. Predicts the field from **weighted slice features**



Application

There are dozens of new architectures published monthly

How do we know which ones are helpful?

Benchmarks

The variety of implementations makes benchmarking difficult

Table 2. Comparison on large geometries benchmarks. Relative L2 of the surrounding area (*Volume*) and surface (*Surf*) physics as well as drag and lift coefficient (C_D, C_L) is recorded, along with their coefficient of determination R^2_D and R^2_L. The closer R^2 is to 1, the better.

Model	DrivAerNet++ Full		DrivAerNet++ Surf			AirCraft		
	Volume ↓	Surf ↓	C_D ↓	R^2_L ↑	Surf ↓	C_L ↓	R^2_L ↑	Surf ↓
GraphSAGE (2017)	0.328	0.284	0.282	0.859	0.294	0.040	0.988	0.109
PointNet (2017)	0.285	0.478	0.301	0.831	0.237	0.095	0.982	0.169
Graph U-Net* (2019)	0.241	0.260	0.272	0.876	0.193	0.063	0.953	0.161
MeshGraphNet* (2021)	0.529	0.422	0.260	0.870	0.209	0.038	0.993	0.113
GNO* (2020a)	0.510	0.664	0.252	0.882	0.196	0.031	0.991	0.129
Galerkin* (2021)	0.234	0.274	0.267	0.792	0.235	0.069	0.879	0.118
Geo-FNO* (2022)	0.718	0.892	0.288	0.831	0.291	0.243	0.903	0.395
GINO (2023a)	0.586	0.638	0.323	0.725	0.220	0.047	0.983	0.133
GNOT* (2023)	0.174	0.171	0.158	0.901	0.167	0.033	0.991	0.093
LNO* (2024)	0.180	0.203	0.208	0.855	0.195	0.091	0.992	0.137
3D-GeoCA* (2024)	0.389	0.224	0.205	0.883	0.175	0.022	0.993	0.097
Transolver* (2024)	0.173	0.167	0.061	0.931	0.145	0.037	0.994	0.092
Transolver++ (Ours)	**0.154**	**0.146**	**0.036**	**0.997**	**0.110**	**0.014**	**0.999**	**0.064**
Relative Promotion	11.0%	12.6%	41.0%	-	24.1%	36.3%	-	30.4%

* These models cannot directly handle million-scale meshes as the model input. Thus, to enable comparison, we split the input mesh
of

Model	Structured Mesh			Point Cloud
	Plasticity	Pipe	Airfoil	Elasticity
WMT (2021)	0.0076	0.0077	0.0075	0.0359
U-FNO (2022)	0.0039	0.0056	0.0269	0.0239
Geo-FNO (2023a)	0.0074	0.0067	0.0138	0.0229
U-NO (2023)	0.0034	0.0100	0.0078	0.0258
F-FNO (2023)	0.0047	0.0070	0.0078	0.0263
LSM (2023)	0.0025	0.0050	0.0059	0.0218
Galerkin (2021)	0.0120	0.0098	0.0118	0.0240
HT-Net (2024)	0.0333	0.0059	0.0065	/
OFormer (2023c)	0.0017	0.0168	0.0183	0.0183
GNOT (2023)	0.0336	0.0047	0.0076	0.0086
FactFormer (2023d)	0.0312	0.0060	0.0071	/
ONO (2024)	0.0048	0.0052	0.0061	0.0118
IPOT (2024)	0.0033	/	0.0088	0.0156
Transolver (2024)	0.0013	0.0043	0.0053	0.0079
SAOT (2025)	0.0014	0.0062	0.0051	0.0090
PGOT (Ours)	**0.0012**	**0.0039**	**0.0046**	**0.0069**
Improvement	7.7%	9.3%	9.8%	12.7%

Model	Point Cloud
	Elasticity
WMT	0.0359
U-FNO	0.0239
Geo-FNO	0.0229
U-NO	0.0258
F-FNO	0.0263
LSM	0.0218
Galerkin	0.0240
HT-Net	/
OFormer	0.0183
GNOT	0.0086
FactFormer	/
ONO	0.0118
Transolver*	0.0060
GFocal (Ours)	**0.0043**
Relative Promotion	28.3%

Table 4. Relative L2 errors (in %) of surface pressure p_s and skin friction coefficient C_f on the NASA-
p_s, volume velocity $\boldsymbol{u}$, wall shear stress τ and volume pressure p_v on the AhmedML and DrivAerMl

Models	NASA-CRM		AhmedML				
	p_s	C_f	p_s	$\boldsymbol{u}$	τ	p_v	p_s
Graph U-Net*	15.85	15.61	6.46	4.15	7.29	5.18	
GINO*	12.39	11.51	7.90	6.23	8.18	8.80	
GAOT*	30.38	59.79	8.02	7.43	9.92	10.47	
UPT	12.78	23.78	4.25	2.73	5.80	3.10	
AB-UPT	9.77	6.43	3.97	1.94	5.60	**2.07**	
Transolver*	9.61	7.04	3.20	1.81	4.85	2.41	
Transolver++*	9.51	6.95	3.47	1.78	5.06	2.35	
Transolver-3	**8.71**	**5.85**	**2.96**	**1.60**	**4.81**	2.16	

Model	RMSE-1	RMSE-all	#Param.	MPNN	Trans.	Hierarchy-free	Coarsen-free
MGN [18]	0.25	15.1	2.0M	✓	×	✓	✓
BSMS-GNN [32]	0.29	16.0	2.8M	✓	×	×	Prior
EvoMesh [35]	0.28	12.9	3.2M	✓	×	×	Learnt
HCMT [23]	–	7.3	2.5M	×	✓	×	Learnt
M4GN [37]	0.26	**2.6**	2.0M*	✓	✓	×	Prior
MGN-T (ours)	**0.10**	3.2	**0.5M**	✓	✓	✓	✓

Benchmarks

A few more independent benchmarks exist:

Table 3: Test errors (%) of branch-trunk neural operators. Bold indica

Model	Heat sink	Bracket	Bracket-time	JEB
DeepONet	0.15%	5.90%	14.8%	-
Geom-DeepONet	0.18%	1.84%	15.1%	-
S-DeepONet	-	-	8.6%	-
S-NOT	-	-	**5.6%**	-
DCON	**0.10%**	**1.75%**	10.7%	-
GANO	0.16%	1.86%	12.9%	**47.1%**

Table 4: Test errors (%) of general-geometry models without parametric inputs.

Model	Heat sink	Bracket	JEB	DrivAerNet	DrivAer++
Best branch-trunk	**0.10% (DCON)**	**1.75% (DCON)**	47.1% (GANO)	23.1% (GANO)	21.8% (GANO)
GNO	5.69%	10.14%	61.1%	45.8%	33.8%
EA-GNO	6.12%	10.88%	58.6%	48.7%	35.8%
MechGraphNet	5.94%	11.27%	62.9%	49.6%	46.5%
GI-FNO	1.21%	3.94%	31.2%	25.6%	19.8%
FigConvUNet	0.89%	3.80%	**29.8%**	23.7%	18.3%
PointNet	6.14%	8.75%	40.1%	28.5%	18.8%
GNOT	5.55%	3.58%	37.6%	17.9%	18.3%
Transolver	5.23%	2.74%	36.6%	**16.7%**	**17.3%**

Zhong, W., Liu, Q., Abueidda, D., Koric, S., & Meidani, H. (2025). A comprehensive comparison of neural operators for 3D industry-scale engineering designs. *ArXiv.org*. https://doi.org/10.48550/arxiv.2510.05995

Benchmarks

NAFEMS Benchmark (results in october):

- AirfRANS Dataset (1000 NACA airfoils)
- Open-Source Models
- Vendor software?

What Works

YASAI

16 models tested in the weak baselines paper were truly faster than numerical solvers

Deep Learning

Plain Neural Networks:
Multilayer Perceptrons
Convolutional Networks (CNN)
Recurrent Networks (RNN)

Graph Neural Networks:
GraphSAGE
GCN
GAT

Transformers:
GPT
Transolver

Neural Operators:
DeepONet
FNO

Successful architectures:

- Plain neural networks
- Convolutional networks
- Recurrent networks

What Works

Common applications

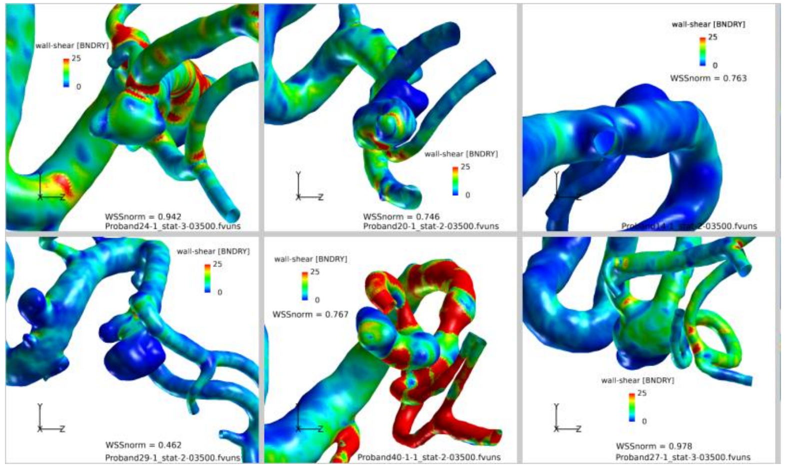
Design Space Exploration:

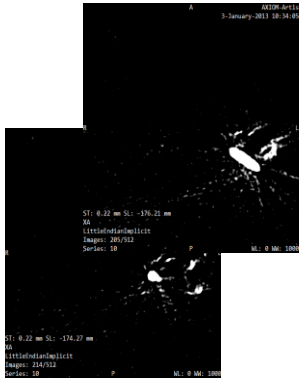
- Efficient DoE
- Optimize Designs
- Often: Tabular Regression

(Near) Real-Time Fields:

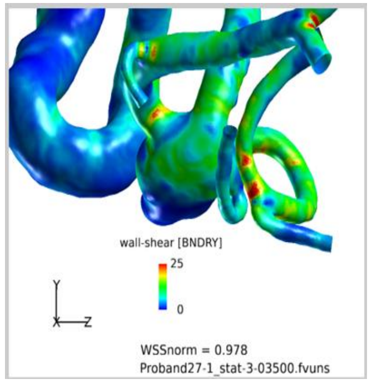
- Frontloaded Simulations
- Useful when time matters
- Essentially: Very efficient lookup and intelligent interpolation

Aneurysm Prediction



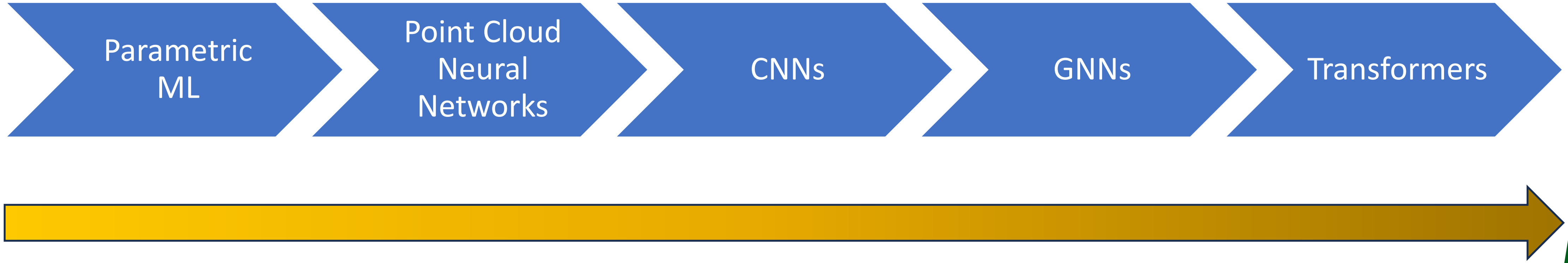


AI Prediction



A larger dataset is used to train the AI

Efficiency



Low Complexity

Megabytes of Data
Datatables
Local CPU Training

Model Complexity ~
(Dataset Size / Inductive Bias)

High Complexity

Petabytes of Data
Lakes of Sim Data
Cloud GPU Training

Pitfalls

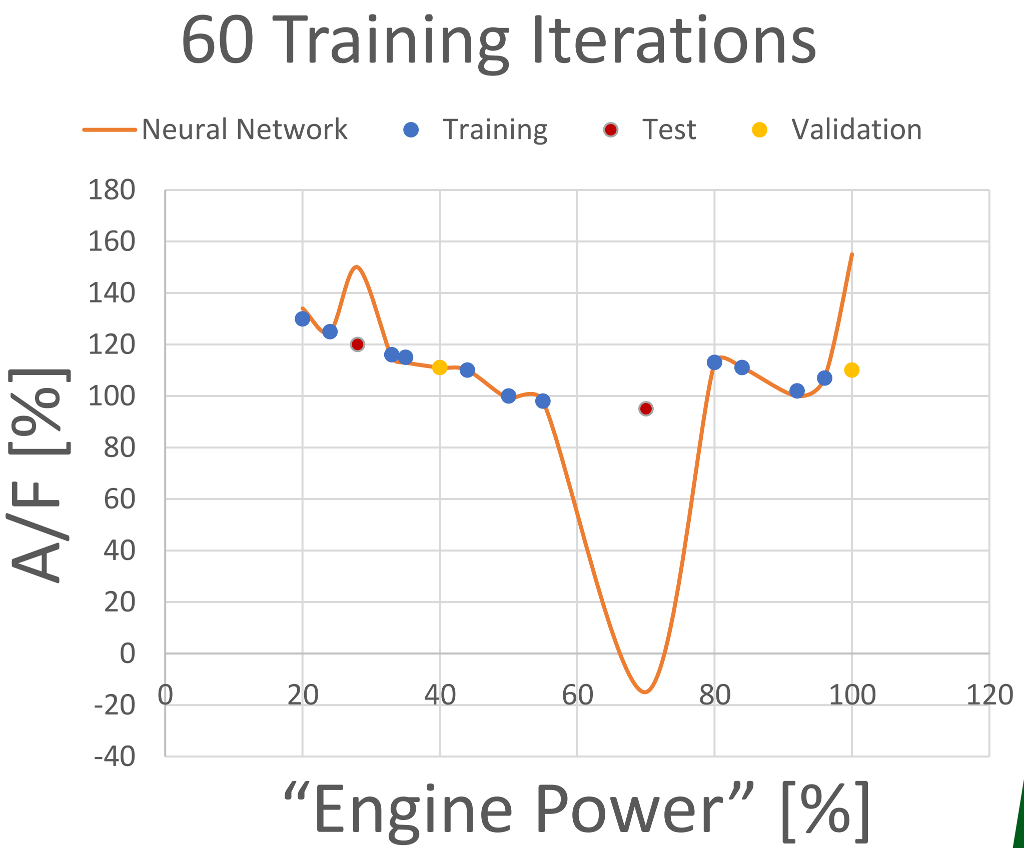
Avoiding pitfalls

Overfitting

Leakage

Artifacts

Overfitting



Leakage

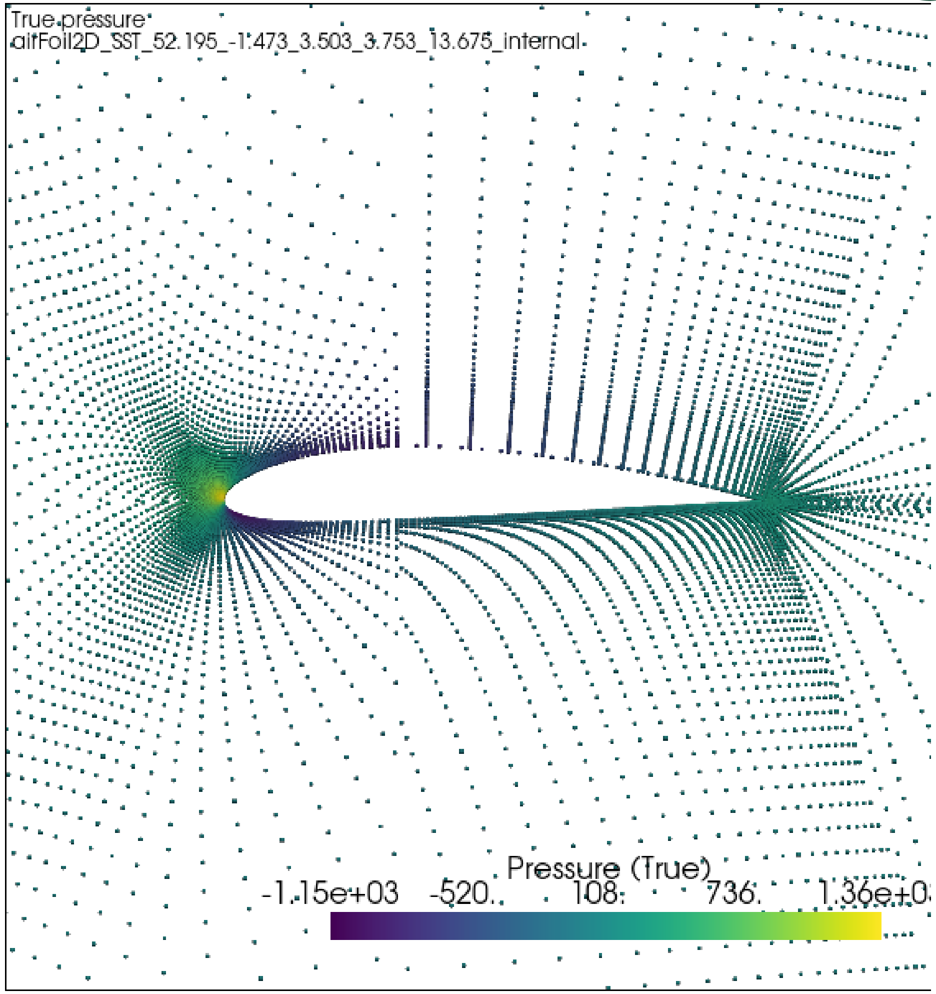
A model can look good on test data

and still be wrong

Leakage

Example:

- Training pressure fields for airfoils
  - Training data contains all geometries
  - Test data uses different held out inlet velocities
- The model will memorize pressure fields



➡ **Predictions for new geometries will be off**

Leakage

Leakage:

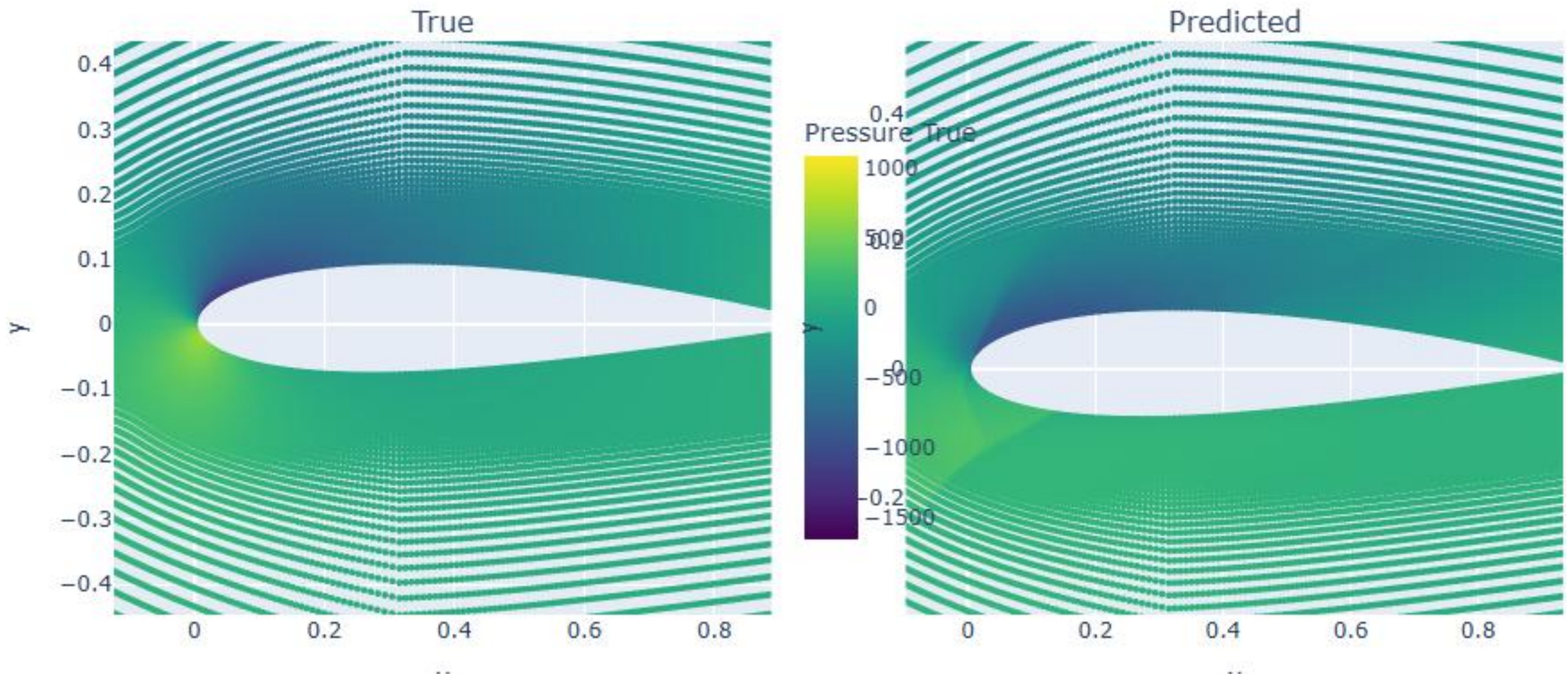
- Model has information at training time that is not available on unseen data
- Analogy: Memorizing without understanding

Effect:

- Models show very good results on test data
- But perform poorly on anything new

Artifacts

Artifacts are hardly quantifiable



PINNs

Physics informed neural networks (PINN):

- Don't have to stick to their physics terms
- Introduce additional numerical instability and computational overhead
- Don't outperform numerical solvers

Physics information can be helpful *and* harmful

Naser, M. (2025). Fundamental flaws of physics-informed neural networks and explainability methods in engineering systems. *Computers & Industrial Engineering*, *212*, 111704. https://doi.org/10.1016/j.cie.2025.111704

More Practical Advice

Prerequisites to build a Sim-AI model:

- Simulation and application knowledge
  - Example Aneurysm AI:
    - Simulation Specialists
    - Radiologists
    - Neurosurgeons
- Cross-sectional ML expertise
- A clear project goal!

More Practical Advice

The most useful AI systems are tailored to a specific application.

There is no general-purpose pipeline.

More Practical Advice

How to tell if an AI product is useful:

Remove AI from the description.

More Practical Advice

YASAI

AI is data-driven by definition:

- It makes use of simulations

- It can’t replace simulation

YASAI

Max Kassera

kassera@yas-ai.com

+49 160 1722246

www.yas-ai.com