Thoughts on The PDE Transformer

by Paul Thompson, 7/2/2025

I suspect that this remarkable new paper [1] on the PDE-Transformer\* will help vision-language models (VLMs) and vision-language-action models (VLAs) to model the external world much better, especially when trained on real-time and offline video.   
Remember that one criticism of large-language models as a means to solve very general problems (AGI/artificial general intelligence) is based on their limited "world model" if only language is used; but this type of method should lead to highly compact (low-MDL) world models that AI methods can use to interpret inputs, think, simulate, and re-think...

Can Transformers Discover Physical Laws?  
I’ve been digging into yesterday's cool new PDE-Transformer paper [1], which learns to discover physical laws (eg how fluid flows) directly from data. But what if multiple physical processes are superimposed - could it still figure things out with no explicit equations?

Multi-channel attention for physics  
These transformer models handle different variables like velocity components u, v and pressure, as separate channels, embedding each into its own tokens. They then apply two attention mechanisms:  
1️⃣ Spatial attention (within each channel): lets the model learn local stencils, approximating derivatives like ∂u/∂x, ∂v/∂x  
2️⃣ Channel attention (across variables at each point): enables it to learn how different quantities interact, eg how pressure gradients drive velocity or how velocity fields transport each other.  
So purely from data, it can discover local derivatives and cross-variable couplings, which is a bit surprising....

Can it learn superimposed PDEs?  
One interesting question: can they discover "subordinated" systems? For example, a Gray-Scott reaction-diffusion field u that’s also advected by a fluid velocity v evolving under Navier–Stokes. Could the model:  
Learn ∂v/∂t​ depends on velocity + pressure?  
Learn ∂u/∂t​ involves local reaction, diffusion, and also advection by v?  
By checking attention maps, you could find out if u queries v (proper advection) while v mostly queries itself + pressure (fluid dynamics).  
So....If it can figure out hidden multi-process dynamics, could it also tackle genetics and multi-omics? learning inter-reacting biological pathways just from data?  
Anyway: With multi-channel spatial + channel-wise attention, transformers can now act as neural PDE solvers. Maybe we should test them on superimposed systems could discover layered dynamics in physics and (maybe?) biology.

A close-up of a computer screen

Description automatically generated

# References

[1] [PDE-Transformer: Efficient and Versatile Transformers for Physics Simulations, Benjamin Holzschuh, Qiang Liu, Georg Kohl, Nils Thuerey, 2025](https://github.com/dimitarpg13/deep_learning_for_dynamical_systems/blob/main/articles/PDE-Transformer-Efficient_and_Versatile_Transformers_for_Physics_Simulations_Holzschuh_2025.pdf)

[2] [Can The PDE Transformer Learn Complex Physical Laws, A Note by Paul Thompson, 7/1/2025](https://github.com/dimitarpg13/deep_learning_for_dynamical_systems/blob/main/docs/Can_the_PDE_Transformer_learn_complex_physical_laws.docx)

[3] [original Linkedin post by Paul Thompson, 7/2/2025](https://www.linkedin.com/posts/paul-thompson-1380216_i-suspect-that-this-remarkable-new-paper-activity-7346041359780204545-GOyo?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAFZfUoBgPoGUucdnvtwuzPv79P8VHj6uvk)