

# Bumblebee: Foundation Model for Particle Physics Discovery

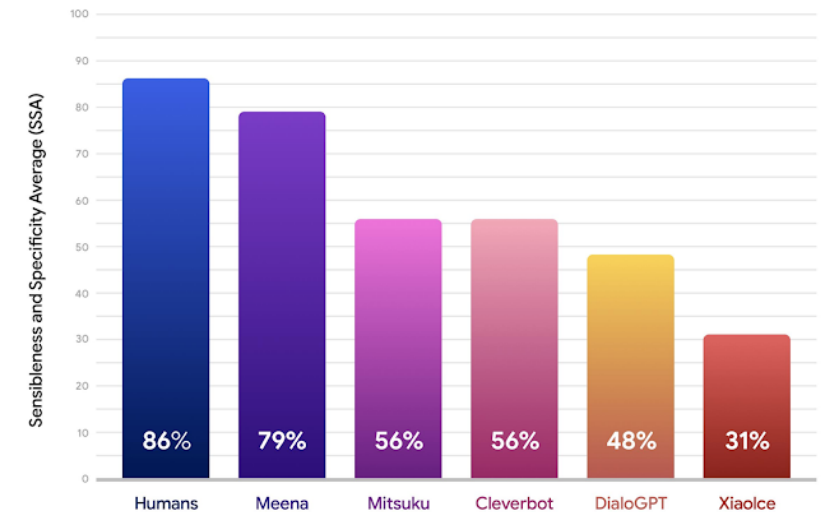
AJ Wildridge, Jack Rodgers, Ethan Colbert,  
Yao Yao, Mia Liu, Andy Jung

2024 US LUA Annual meeting

[arXiv:2412.07867v1](https://arxiv.org/abs/2412.07867v1)

# What are **Foundational** Models?

- **Foundational** models learn representations of their inputs that can be used in many downstream tasks
- Popularized in natural language processing (NLP)
- Enables **state of the art** performance where supervised techniques struggle



[arXiv:2001.09977](https://arxiv.org/abs/2001.09977)

# What are **Foundational** Models?

- **Foundational** models learn representations of their inputs that can be used in many downstream tasks



**Can we create a foundation model to discover new physics?**

- Popular (NLP)

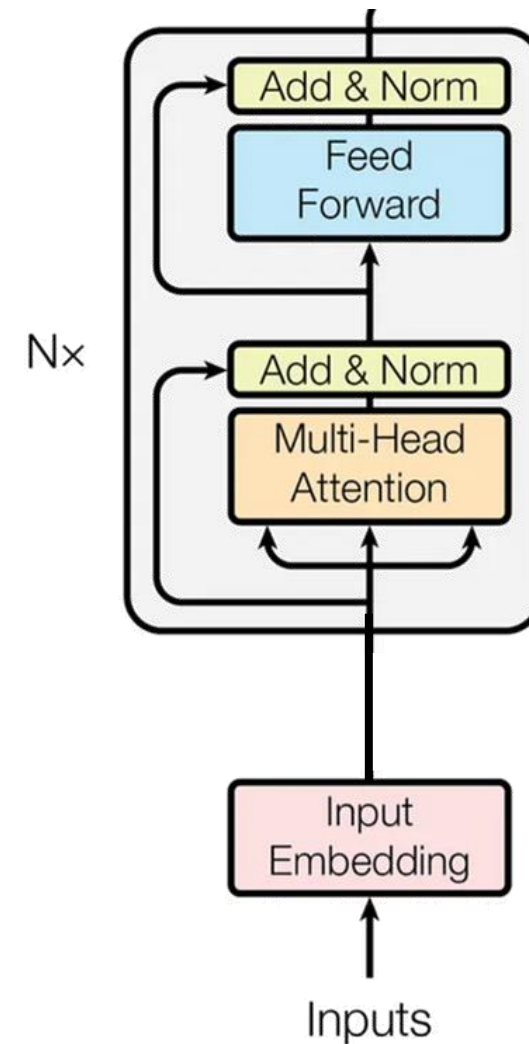
- Enables **state of the art** performance where supervised techniques struggle



# Bumblebee

[arXiv:1810.04805](https://arxiv.org/abs/1810.04805)

- Same spirit as BERT
- Particle physics is **permutation invariant**  
→ remove positional encodings
- Particle vector learned **input embeddings**  
in **Bumblebee**
- 8 encoder layers, 768-dimensional hidden state, 16 self-attention heads  
→ 57M parameters



# Bumblebee – Pre-training

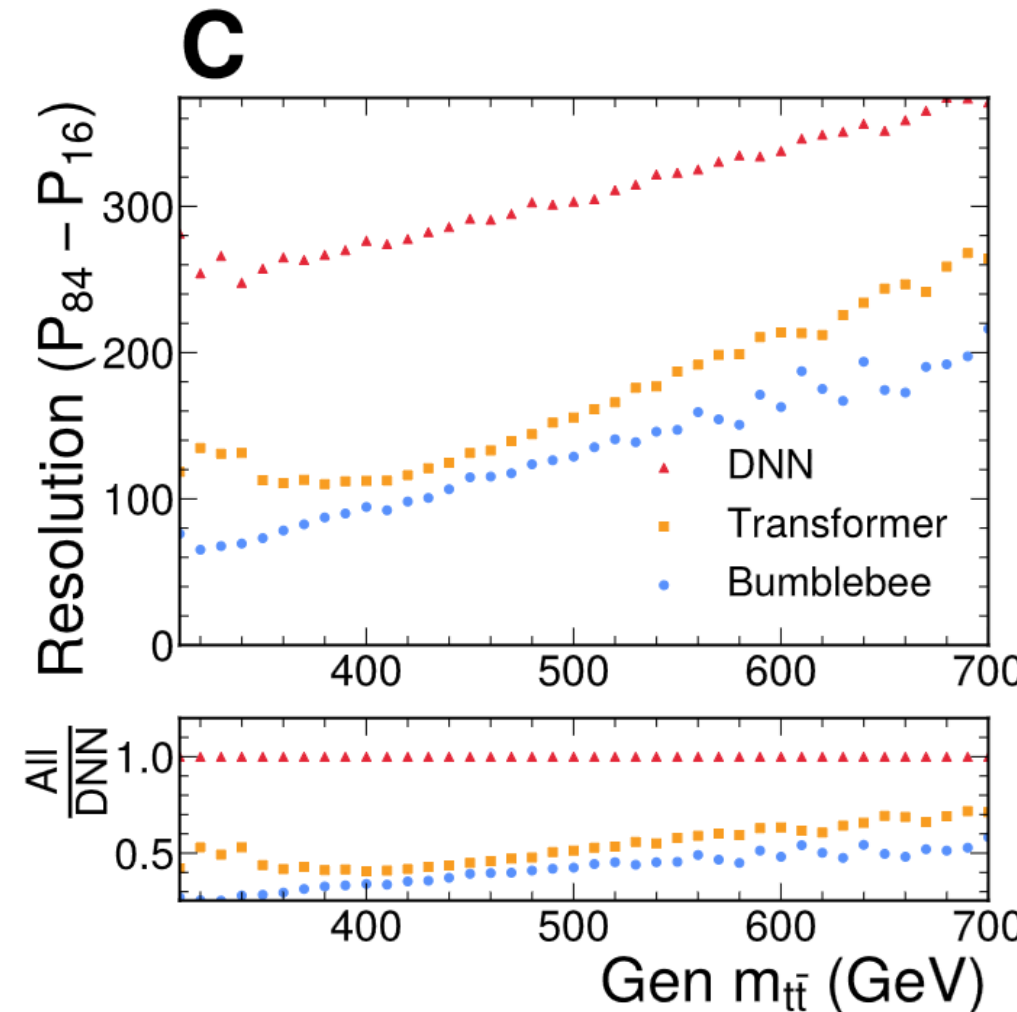
- Masked **pre-training** task
  - 50% of the time randomly mask a particle
  - 25% (25%) of the time mask all reco (gen) particles
- Backpropagate on MSE loss on masked particles

## Input Embedding

Input	Reco. $e^-$	Gen. $\bar{\nu}_\mu$	Reco. $\mu^-$	Gen. $\nu_e$	Gen. $b$	Reco. $e^+$	Gen. $\mu^-$	Gen. $b$	Reco. b-jet	Gen. $e^+$	Reco. light jet	Reco. light jet	Reco. MET
Particle Vector Embeddings	$p_T$	$p_T$	$p_T$	$p_T$	$p_T$	$p_T$	$p_T$	$p_T$	$p_T$	$p_T$	$\vec{0}$	$p_T$	$p_T$
	$\eta$	$\eta$	$\eta$	$\eta$	$\eta$	$\eta$	$\eta$	$\eta$	$\eta$	$\eta$		$\eta$	0
	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$	$\phi$		$\phi$	$\phi$
	$m$	$m$	$m$	$m$	$m$	$m$	$m$	$m$	$m$	$m$		$m$	0
	0	0	0	0	1	0	0	1	b	0		b	0
	+	+	+	+	+	+	+	+	+	+	+	+	+
PDG ID Embeddings	11	-14	13	12	5	-11	13	-5	5	-11	41	41	40
	+	+	+	+	+	+	+	+	+	+	+	+	+
Level Type Embeddings	1	0	1	0	0	1	0	0	1	0	1	1	1
	+	+	+	+	+	+	+	+	+	+	+	+	+
Mask Status Embeddings	0	0	0	0	0	0	0	0	0	0	1	0	0

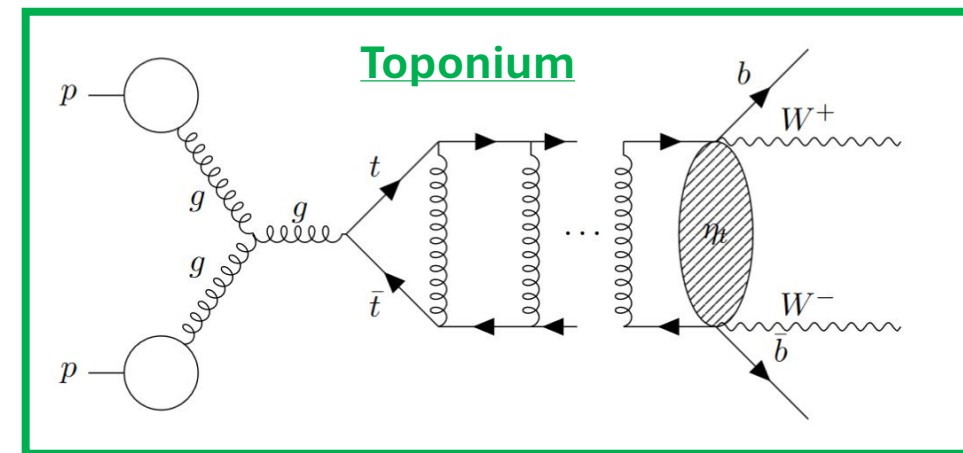
# Bumblebee – Pre-training Results

- **Pretrained** on **dileptonic  $t\bar{t}$**  events
- **Pre-training** task of mask gen and predict gen is akin to neutrino **reconstruction**
- **Bumblebee’s “reconstruction”** outperforms supervised ML techniques

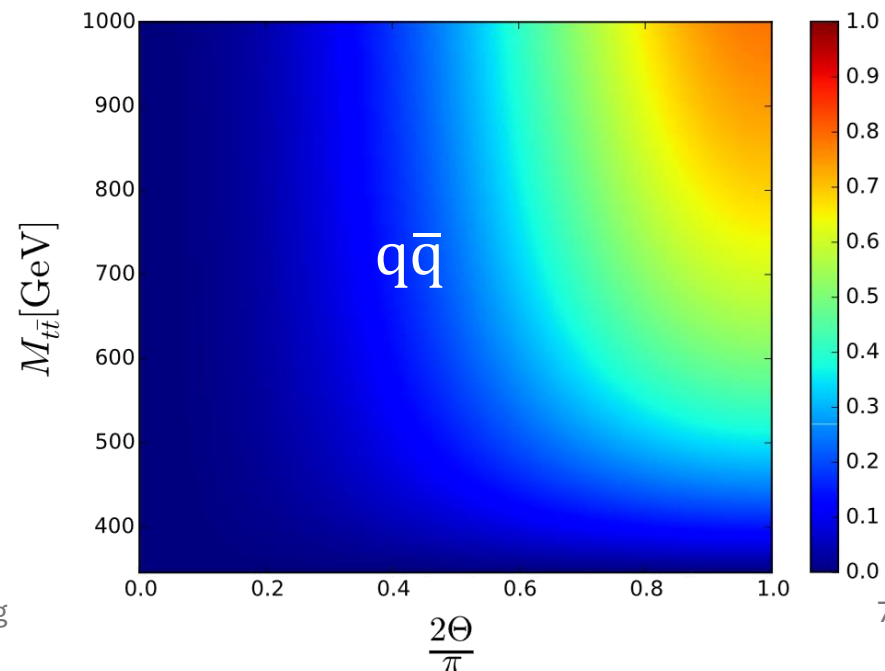
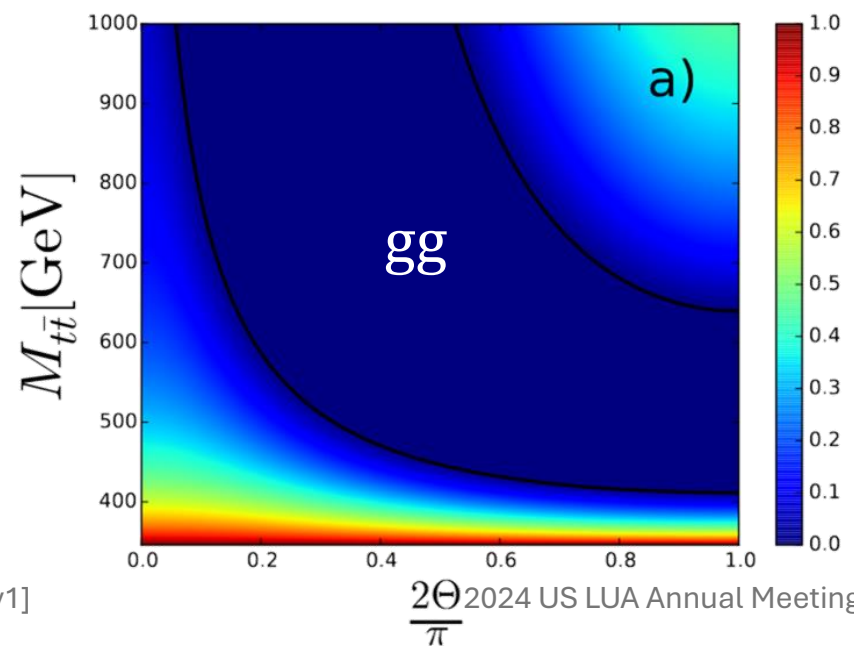


# Bumblebee – Fine-tuning

- Always give reco, mask gen
- Classification tasks: add masked vector  $(1, 0, \dots, 0)/(0, \dots, 0)$  for signal/bkg
  - Fine-tuned on predicting this vector
- We consider two **fine-tuning** tasks:
  - Bound  $t\bar{t}$  state (**toponium**) discrimination
  - $t\bar{t}$  initial state classification



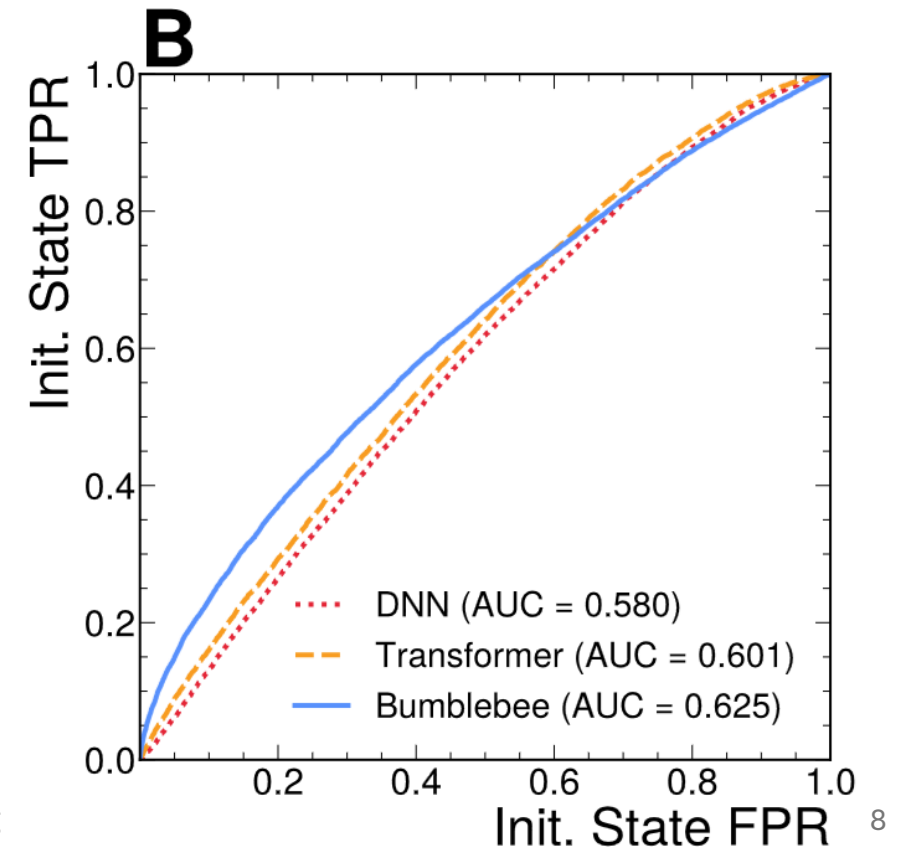
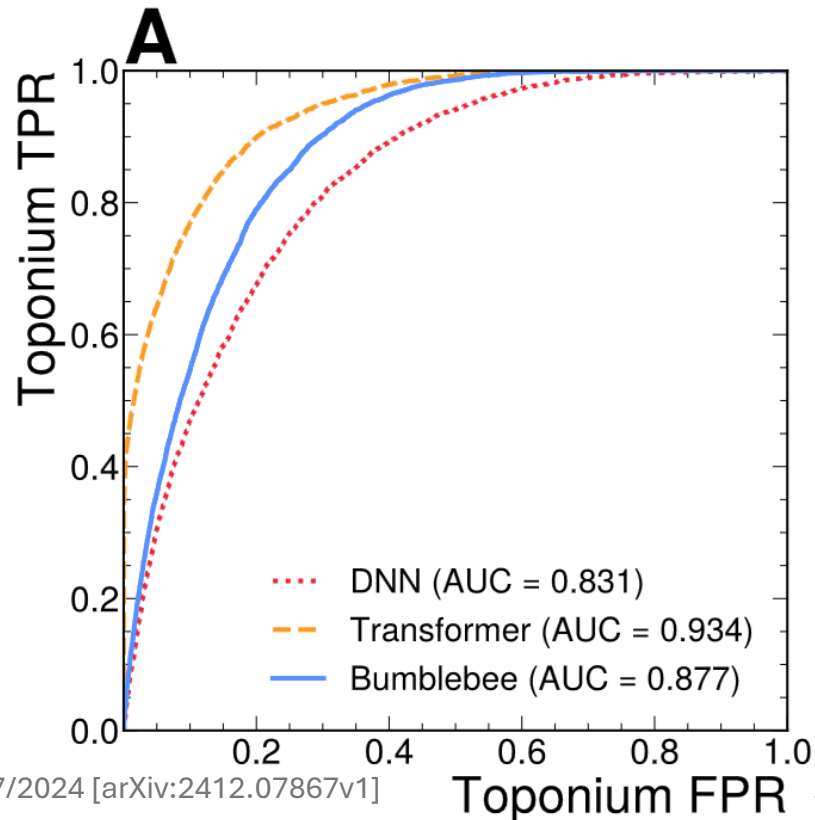
[*Eur. Phys. J. Plus* **136**, 907]



# Bumblebee – Downstream task performance

- Use toy model of ground state of **toponium**,  $\eta_t$
- MLP 0.831 vs Bumblebee 0.877 AUC

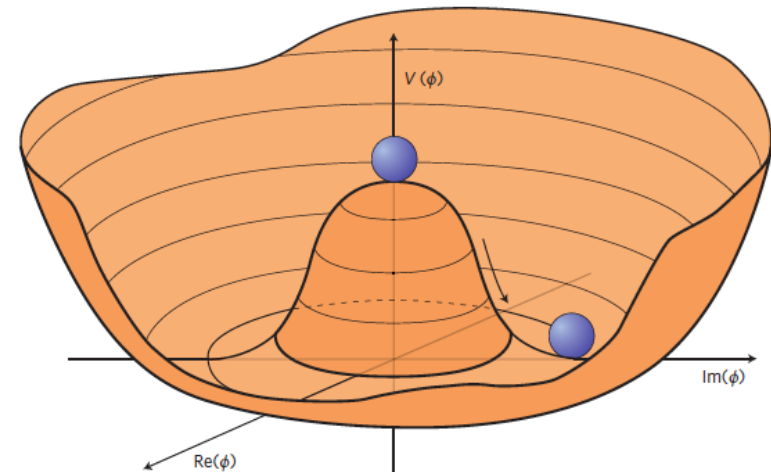
- Bumblebee improves state-of-the-art on a difficult task





# Summary

- **Bumblebee** is a **foundational** model and when fine-tuned on 2 downstream tasks **competitive w/ supervised methods**
- Plenty of exciting use cases for Bumblebee
  - **Easily extend to other event topologies**
- Pre-training objective alone very useful



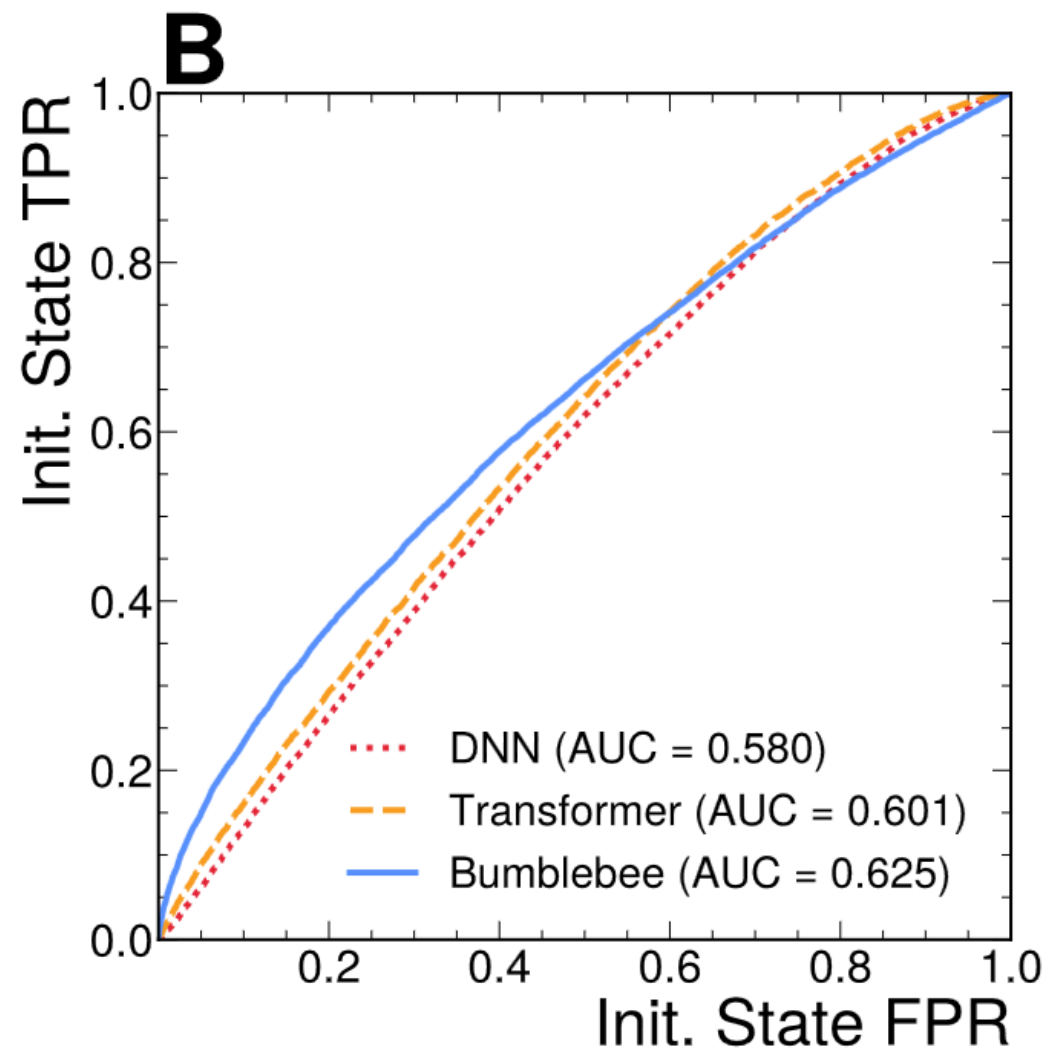


# Thanks!

[awildrid@purdue.edu](mailto:awildrid@purdue.edu)

# Bumblebee – $t\bar{t}$ Initial State Classification

- Initial state is very sensitive to discovering new physics
  - e.g.  $A_C$  in top quarks [Phys. Lett. B 707, 92 \(2012\)](#)
- Also applicable for top quark spin / property measurements
  - gg initial state is spin singlet at production threshold [\[Eur. Phys. J. Plus 136, 907\]](#)
- Bumblebee sees improvement over MLP & Transformer
- **Bumblebee improves state-of-the-art on a difficult task**

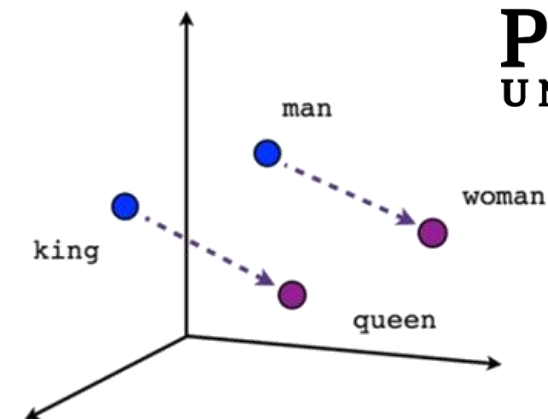


# Bumblebee - Datasets

- 7M  $t\bar{t}$  signal sample produced using POWHEGv2 event generator at next-to-leading-order (NLO)
- 1M  $\eta_t$  signal sample produced using MG5 aMC@NLO generator at LO
- Parton showering and hadronization using Pythia8 in all samples
- Delphes detector simulation
- Trained after object & event selection

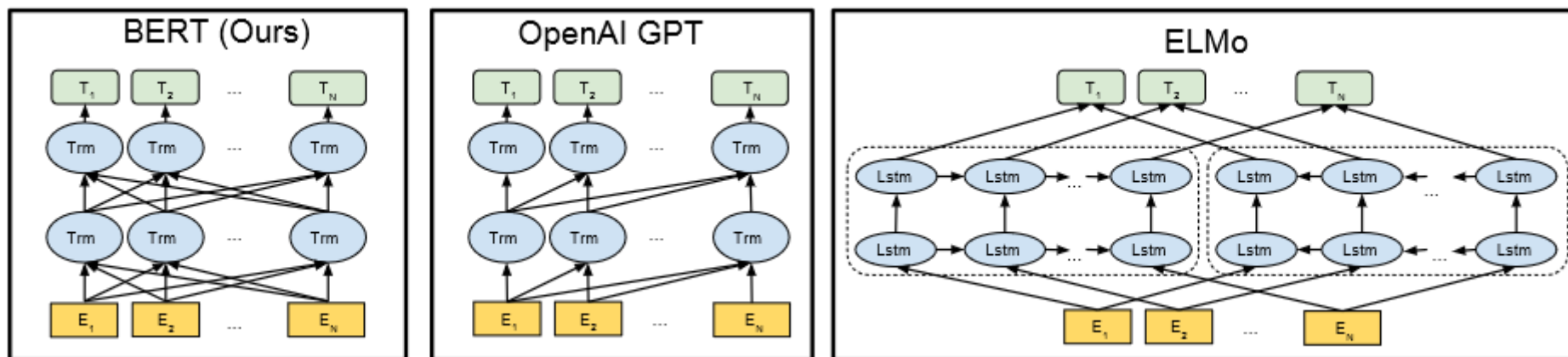
# Foundational Models

- [word2vec](#) & [GloVe](#) created “thought vectors”
  - e.g. (king - man) + woman = queen
  - Powerful learned embeddings for other tasks



Male-Female

- [BERT](#), [GPT](#) and [ELMo](#) moved to DNNs / Transformers for this task



# Bidirectional Encoder Representations from Transformers (**BERT**) [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)

- Used only the **encoder** portion of transformer
- Two tasks: **pre-training** and **fine-tuning**
- **Pre-trained** model learns a vector embedded representation that can be used for downstream tasks

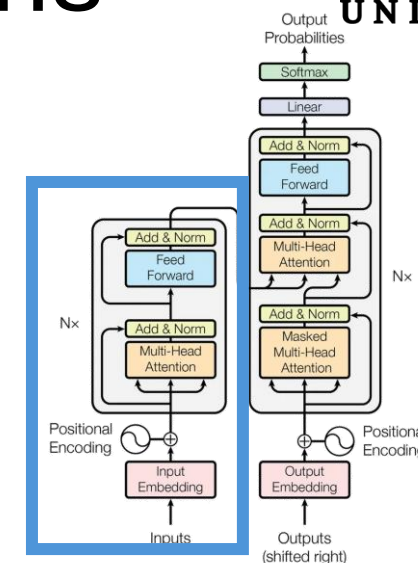


Figure 1: The Transformer - model architecture.

