

Welcome

Triton Conference 2024



Aparna Ramani

VP, Engineering

Trends in Medicine

How Artificial Intelligence is Disrupting Medicine and What it Means for Physicians

AI is going to eliminate way more jobs than anyone realizes



RESEARCH ARTICLE | ✓

AI-assisted superresolution cosmological simulations

The dawn of the omnistar

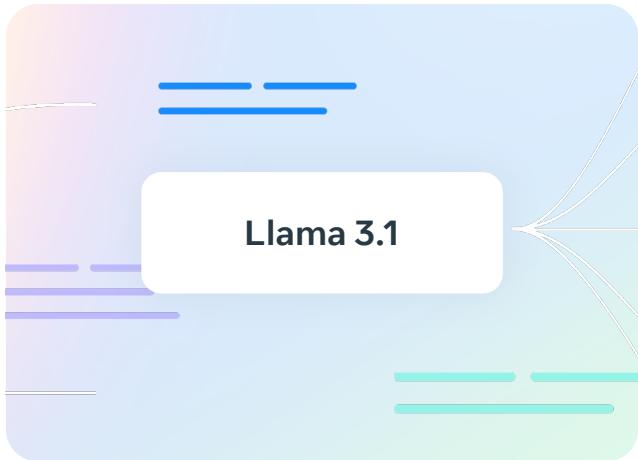
How artificial intelligence will transform fame



COMMENTARY

AI can strengthen U.S. democracy—and weaken it

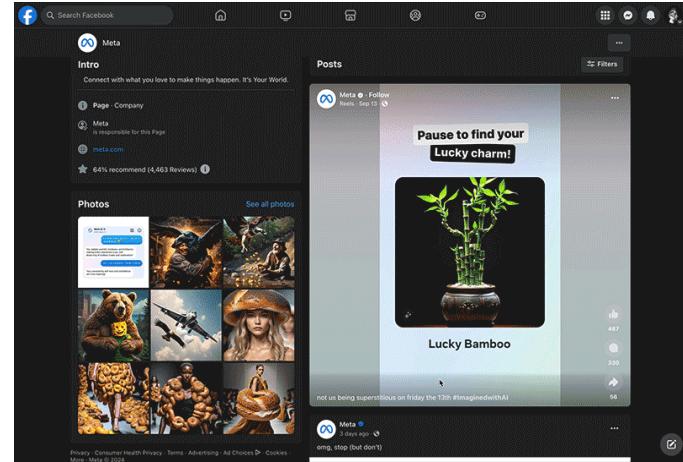
AI's Moment



Large language models
(LLMs)

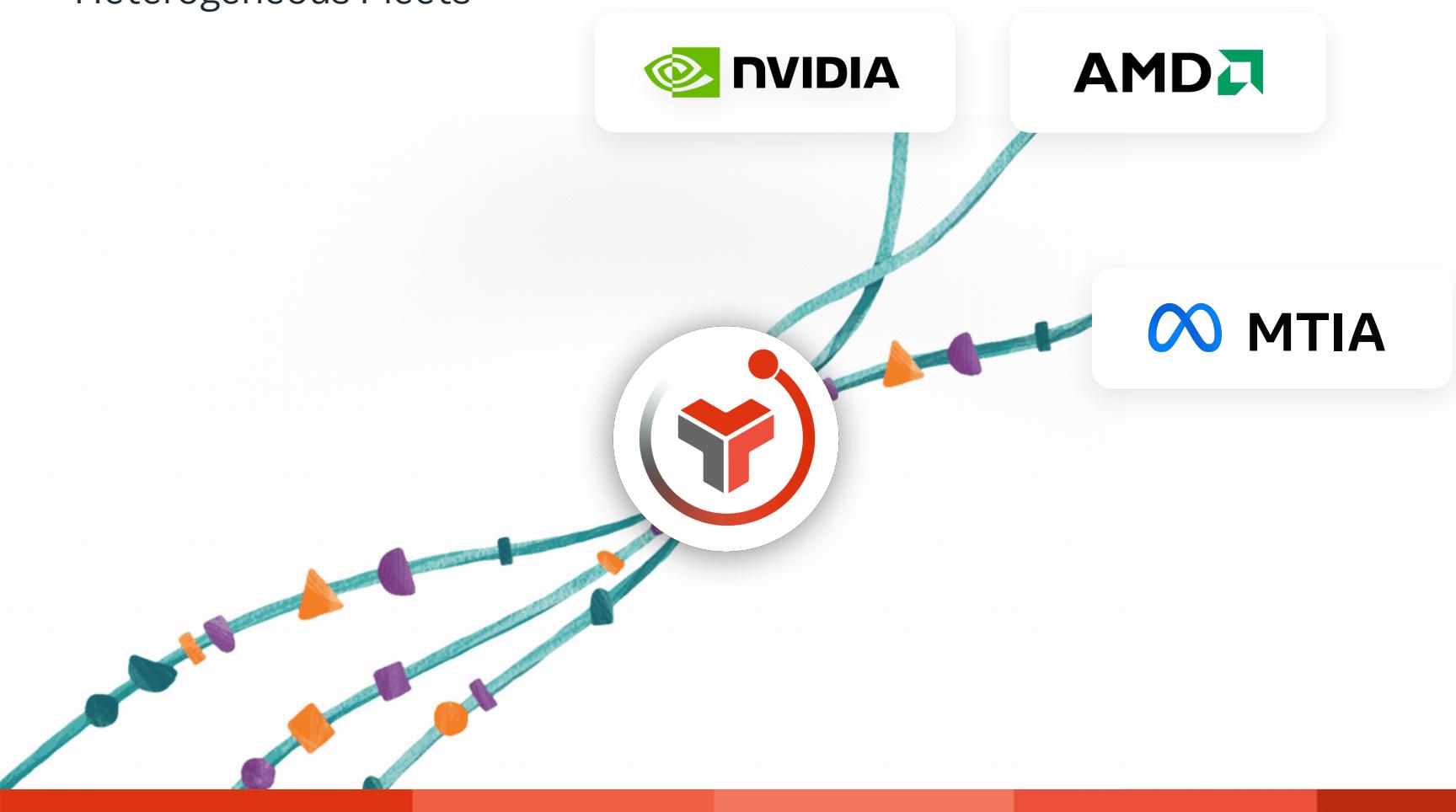


Ads



News Feed

Heterogeneous Fleets





Open models

Distributes innovation

Open Models

Llama 3 8B

April

Llama 3 70B

April

Llama 3.1 405B

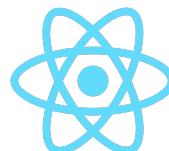
July

Open Source



OPEN
Compute Project®

 PyTorch



React

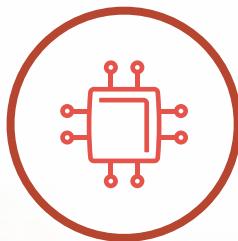




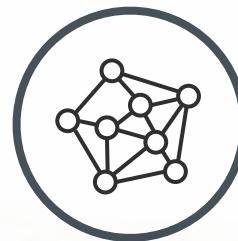
Scaling & Model Architecture



Data



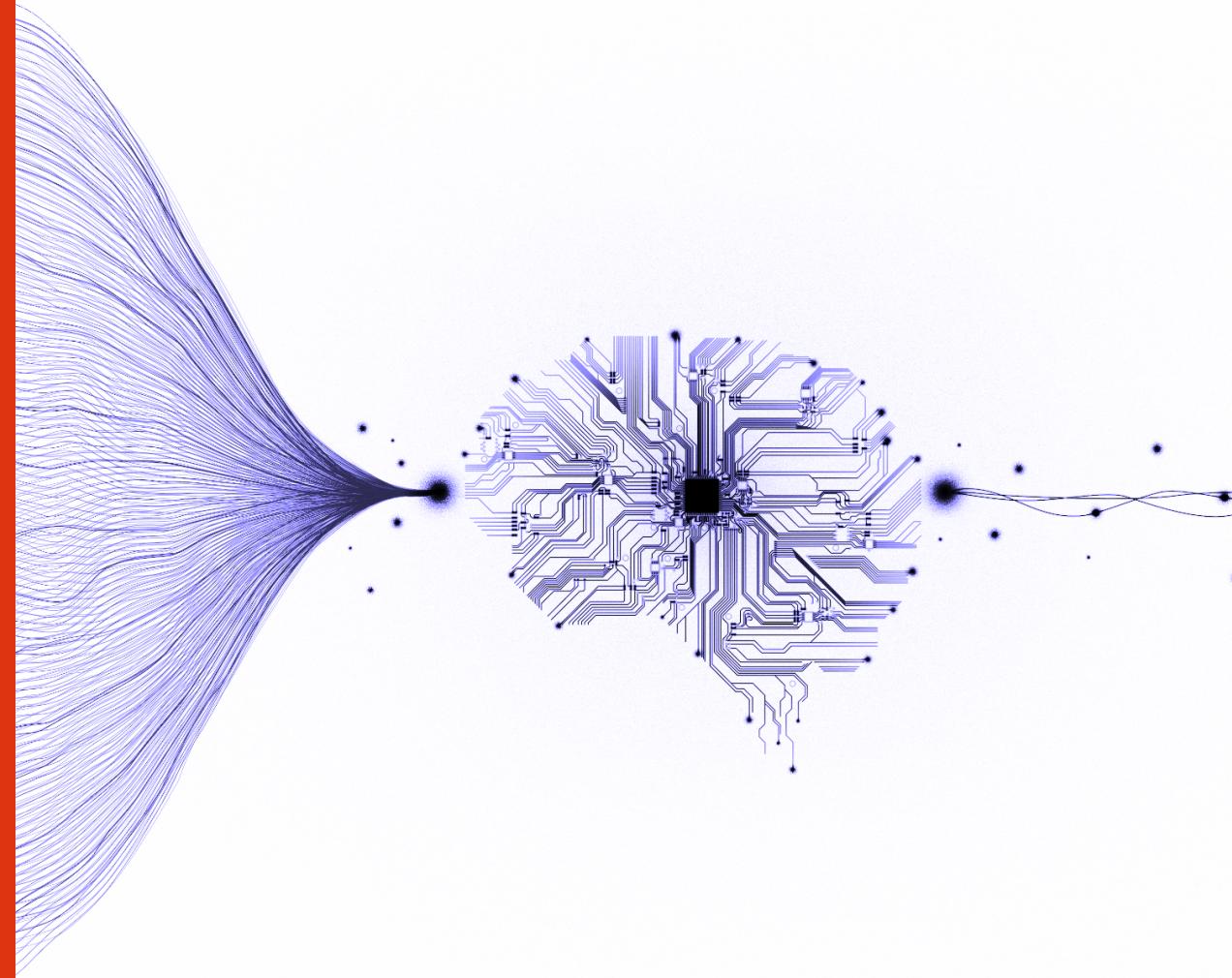
Compute

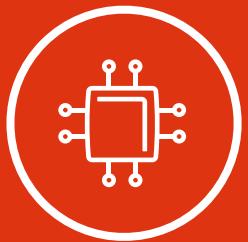


Model
Architecture



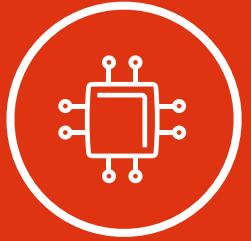
Data



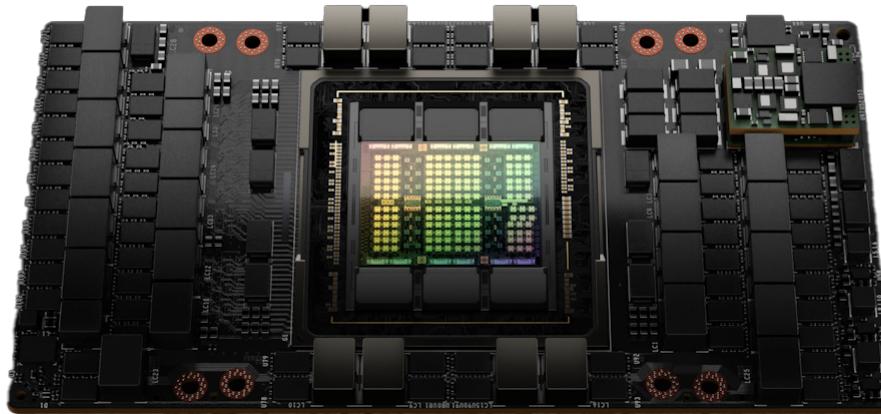


Compute

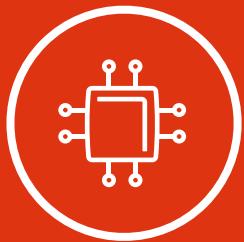




Compute

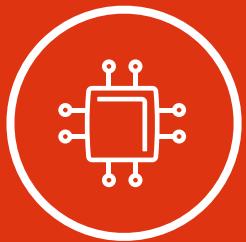


Procuring GPUs



Compute





Compute





Model
Architecture

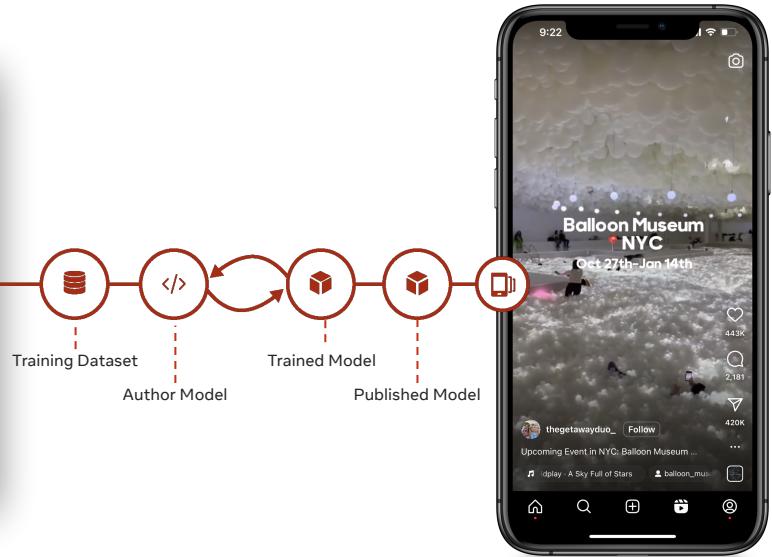
**Generalized Entropy Regularization or:
There's Nothing Special about Label Smoothing**

Clara Meister¹ Elizabeth Salesky³ Ryan Cotterell^{1,2}
¹ETH Zurich ²Johns Hopkins University ³University of Cambridge
 clara.meister@inf.ethz.ch esalesky@jhu.edu
 ryan.cotterell@inf.ethz.ch

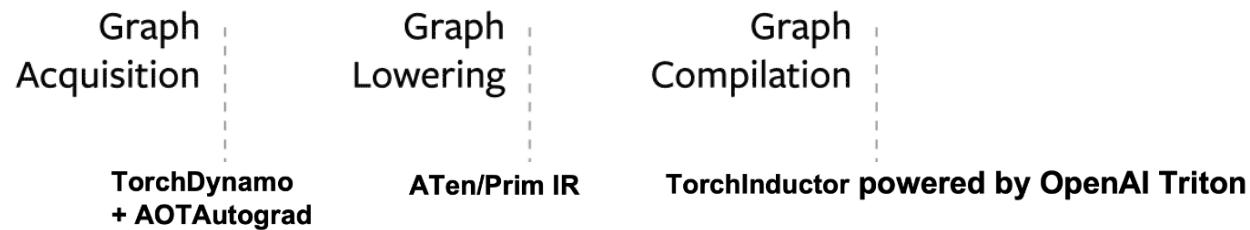
Abstract

Prior work has explored directly regularizing the cross-entropy loss to alleviate over-confidence (i.e. over-confident) predictions, a common sign of overfitting. This class of techniques, of which label smoothing is one, has a connection to entropy regularization. Despite this, there is little understanding of how entropy regularization methods perform across architectures and datasets in language generation tasks, two problems remain open: (1) the specific ways in which entropy regularization affects entropy regularizers have on models, and (2) the full space of entropy regularization techniques is largely unexplored. We introduce a new entropy regularizer, Generalized Entropy Regularization (GER), which includes label smoothing as a special case, and use it to gain a better understanding of the relationship between the performance of a trained model and its performance on language generation tasks. We also find that variance in model performance is often explained by the resulting entropy of the model. Lastly, we find that label smoothing probably does not allow for sparse distributions, an undesirable property for language generation models, and therefore advise the use of other entropy regularization methods in its place. Our code is available online at <https://gitlab.yicoliab.net/entropy regularization>.

1 Introduction



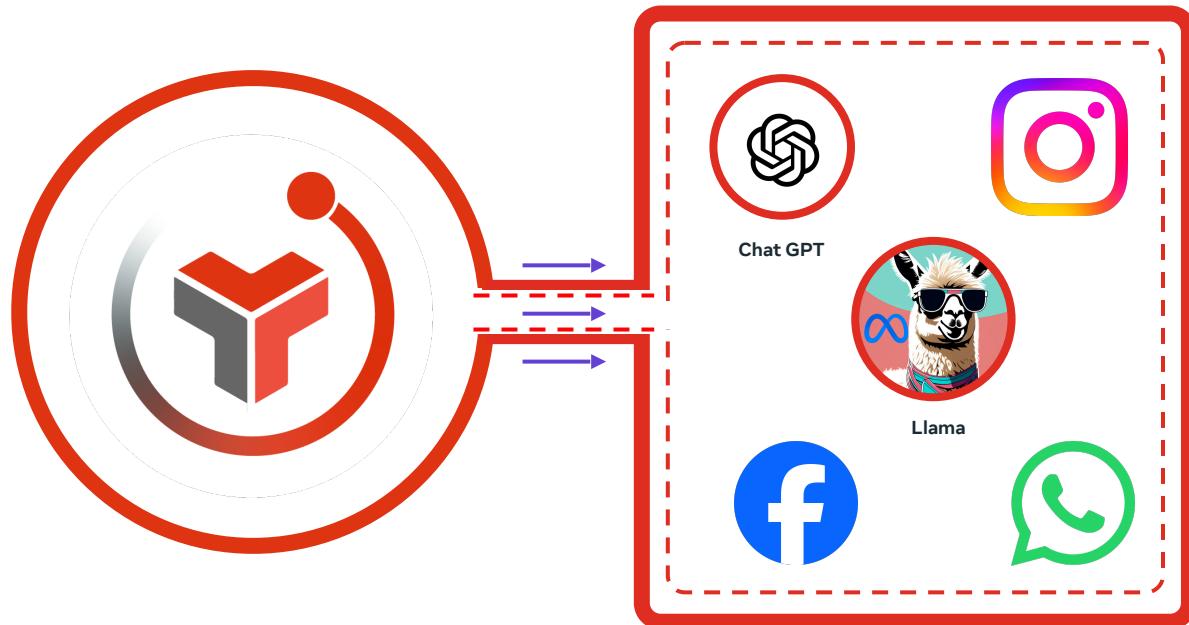
PyTorch Inductor Backend

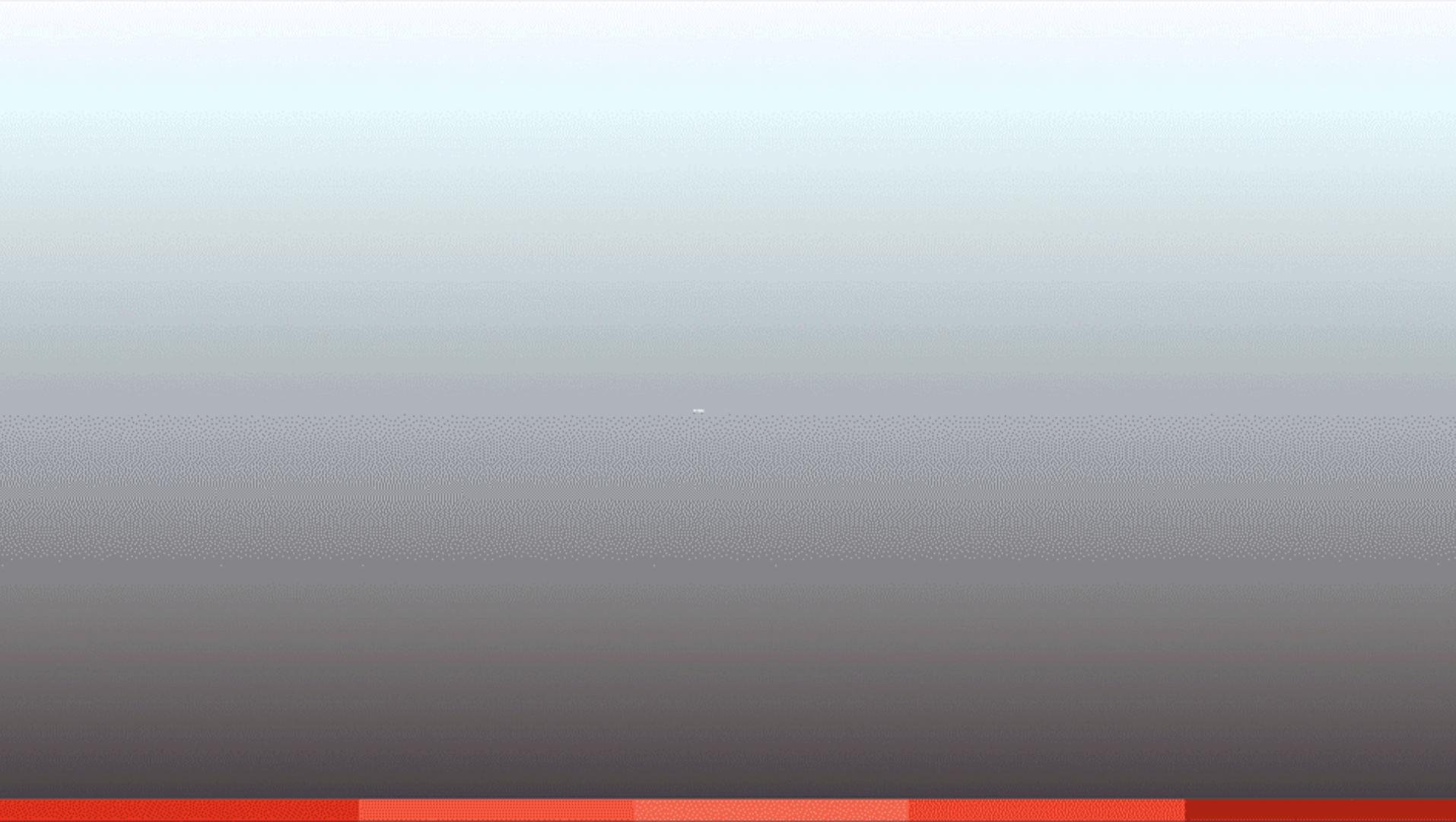


```
def foo(x):
    y = F.conv2d(x, ...)
    z = F.batch_norm2d(y, ...)
    return F.relu(z)
```



Move Fast on Model Architecture Innovation











Thank you