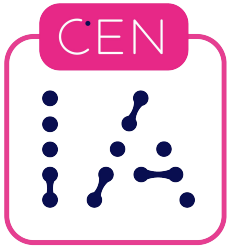
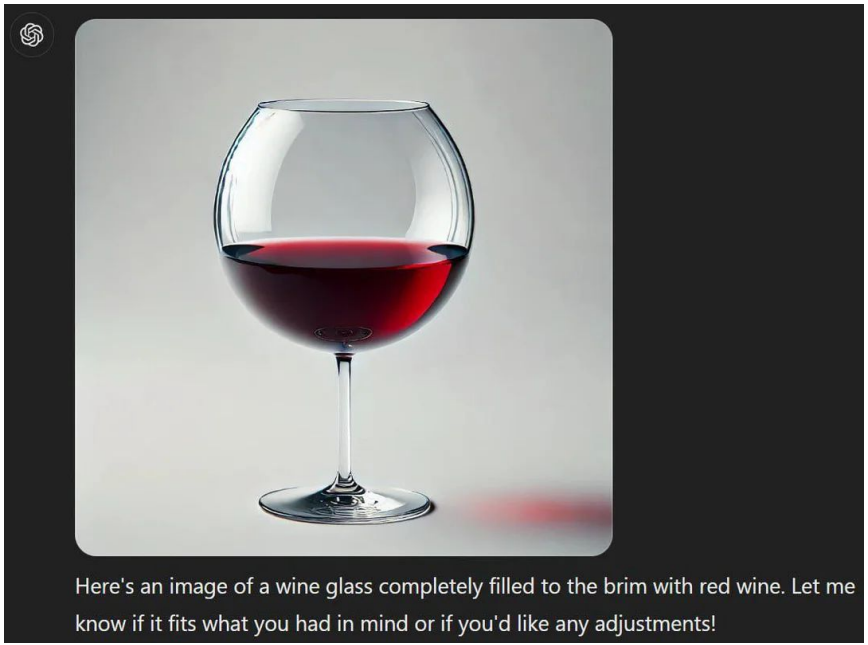






# Data Distributional Properties As Inductive Bias for Systematic Generalization

Felipe del Rio, Alain Raymond, Daniel Florea, Rodrigo Toro Icarte,
Julio Hurtado, Cristian B. Calderon, Alvaro Soto
U. Catolica de Chile, CENIA & U. of Warwick

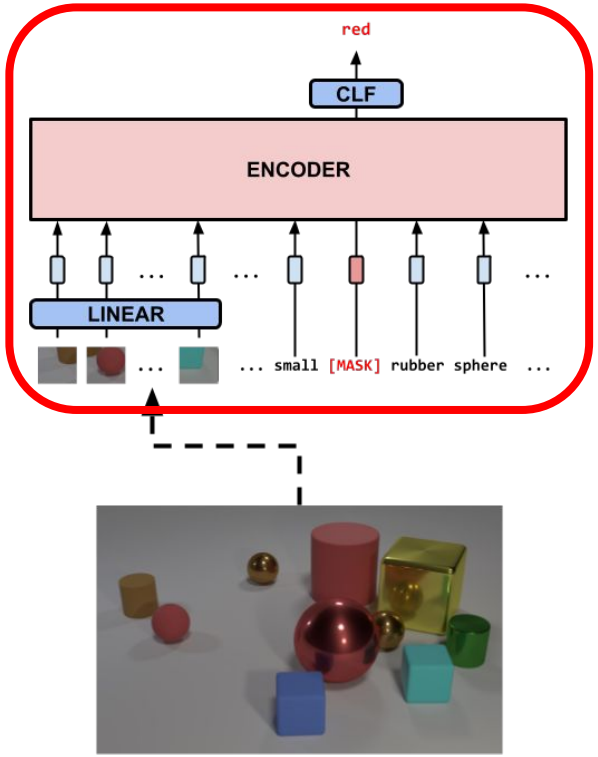
# Models Struggle with Systematic Generalization




**Which data distributional properties help models acquire systematic generalization?**

Image from: https://www.reddit.com/r/ChatGPT/comments/1i3jfrm/chatgpt_cant_do_a_full_glass_of_wine_prove_me/

# Framing the Problem: How We Test Data-Induced SG

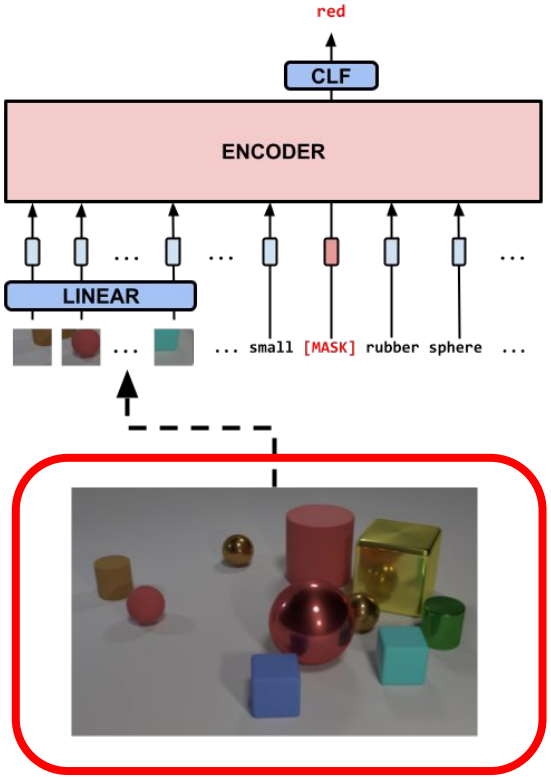



Multimodal Masked Language Model Task

Vision + Text Transformer Encoder

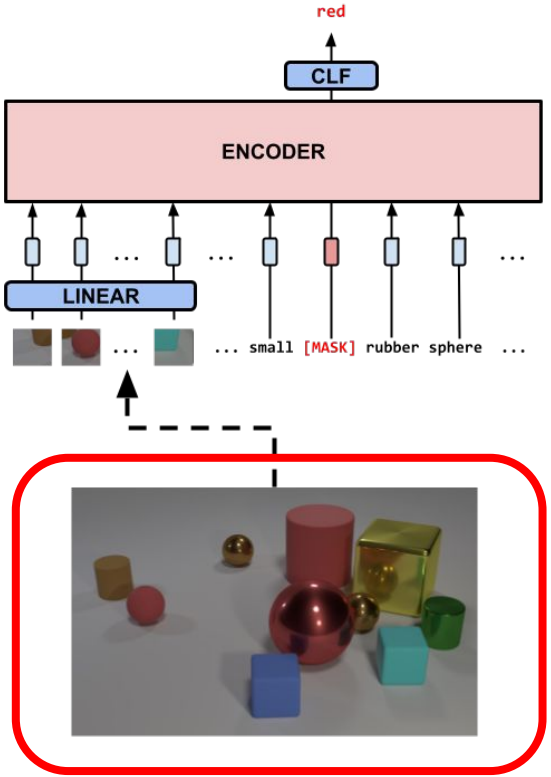
Receives an Image and structured text with masked properties

# Framing the Problem: How We Test Data-Induced SG

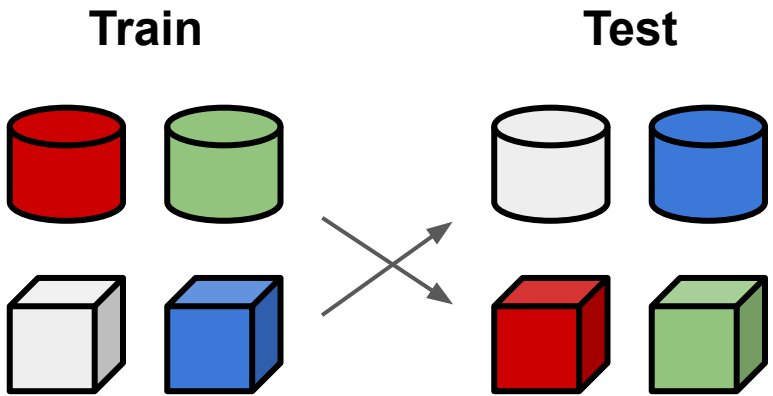



We built on top of the CLEVR CoGenT benchmark

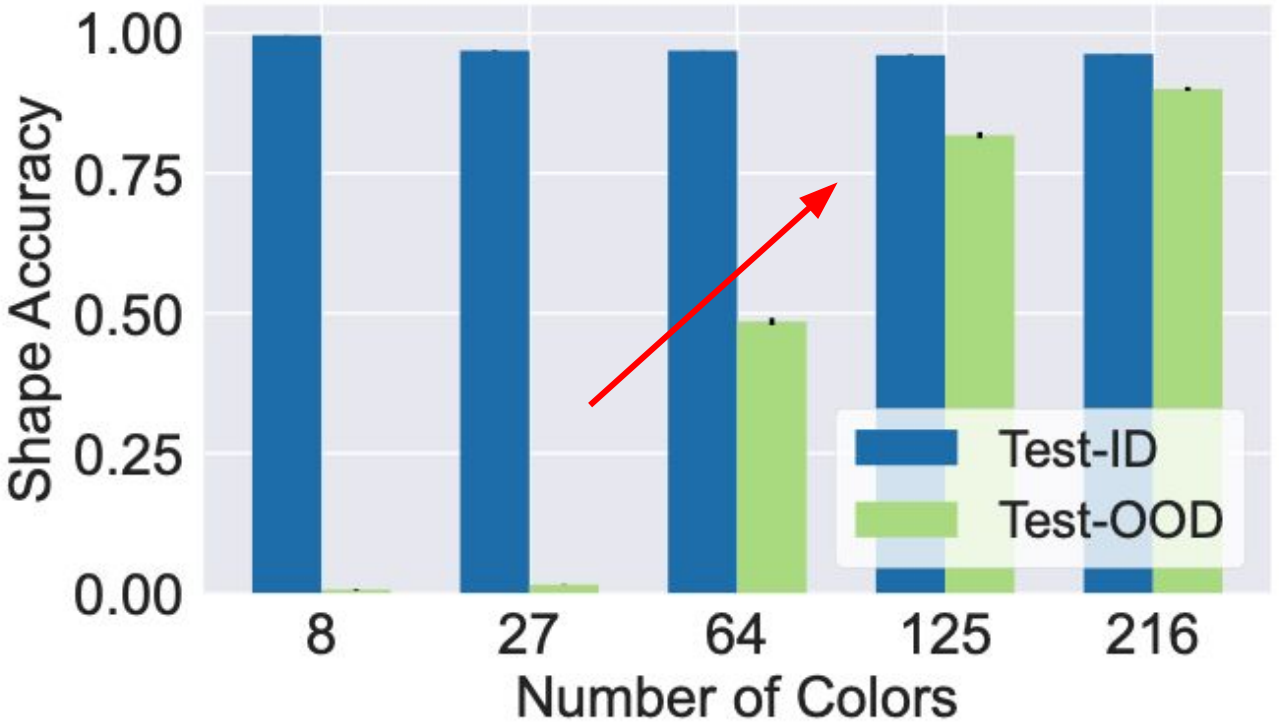
# Framing the Problem: How We Test Data-Induced SG




Several datasets with systematic differences between train and test

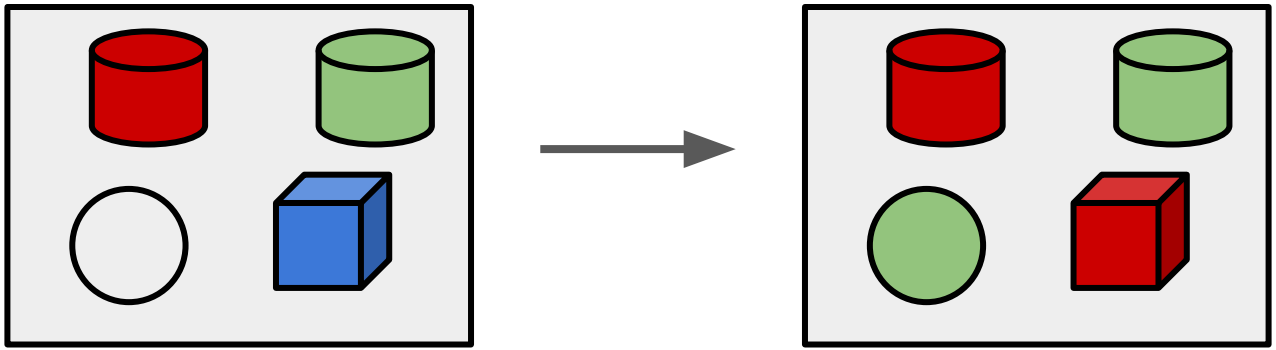


# Diversity: How Color Diversity Boosts Generalization



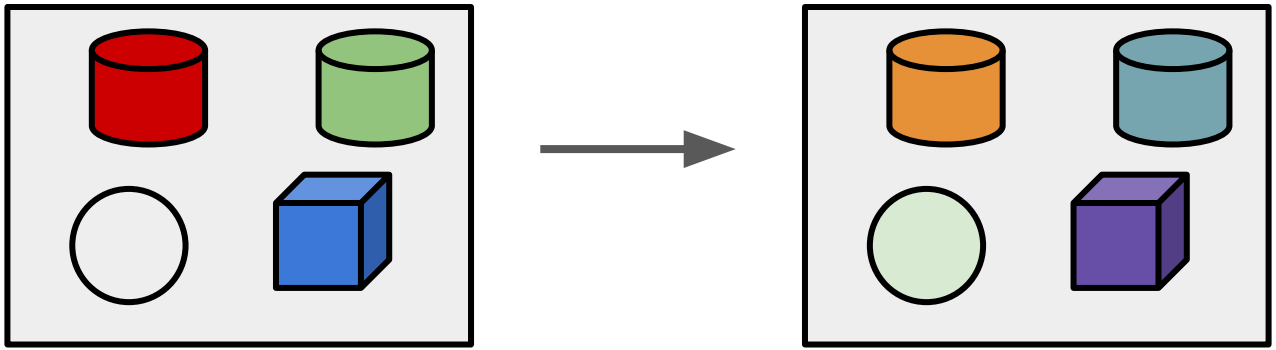
# Burstiness and Latent Intervention: Complementary Gains for Generalization

Burstiness



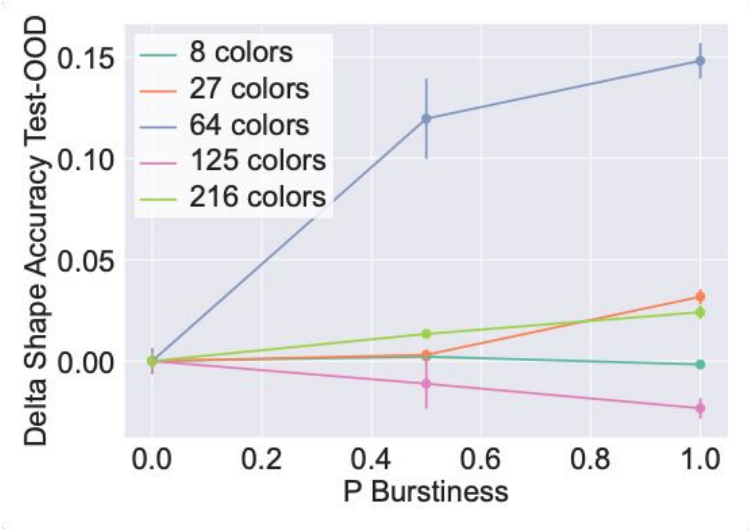
# Burstiness and Latent Intervention: Complementary Gains for Generalization

Latent Intervention

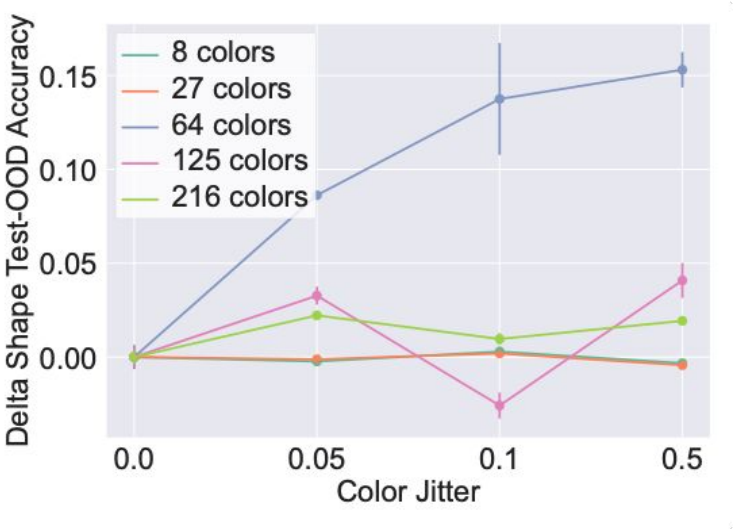


# Burstiness and Latent Intervention: Complementary Gains for Generalization

Burstiness




Latent Intervention

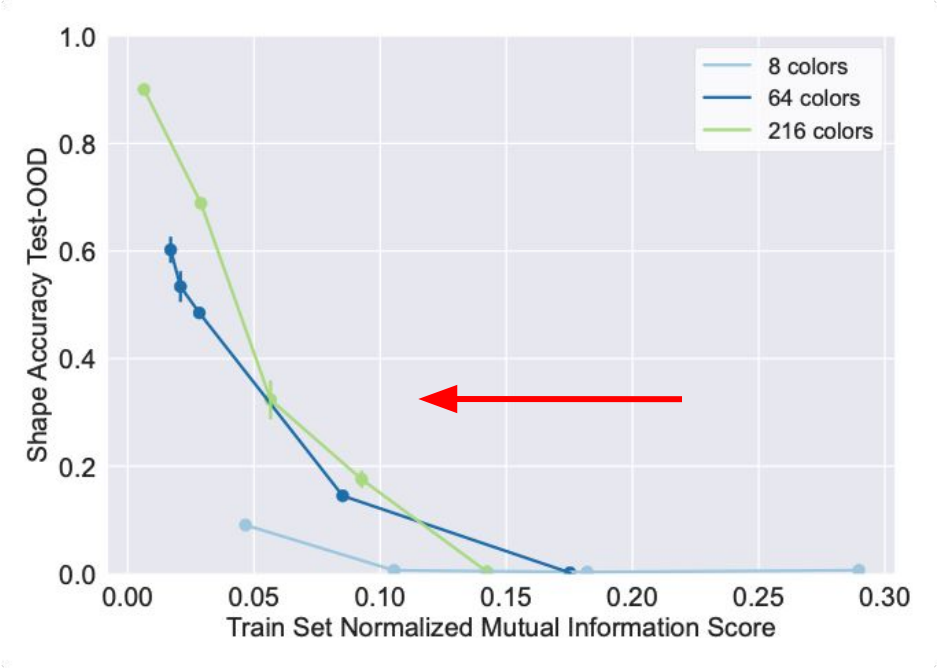


# Why Does Diversity Improve Generalization?

Several hypotheses:

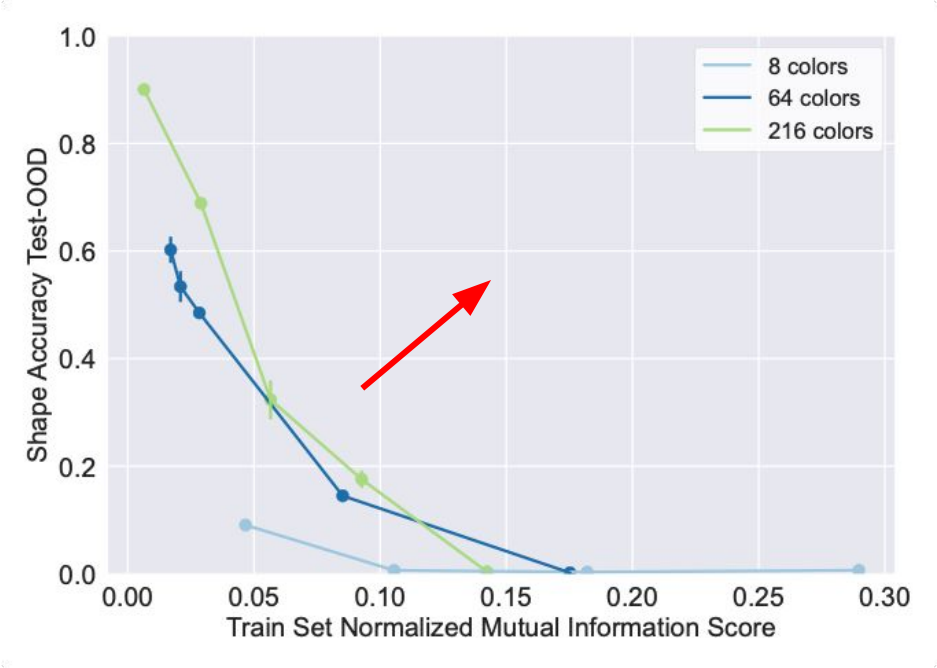
- Model capacity
- Disentanglement
- Reduced shortcut learning

# Why Does Diversity Improve Generalization?



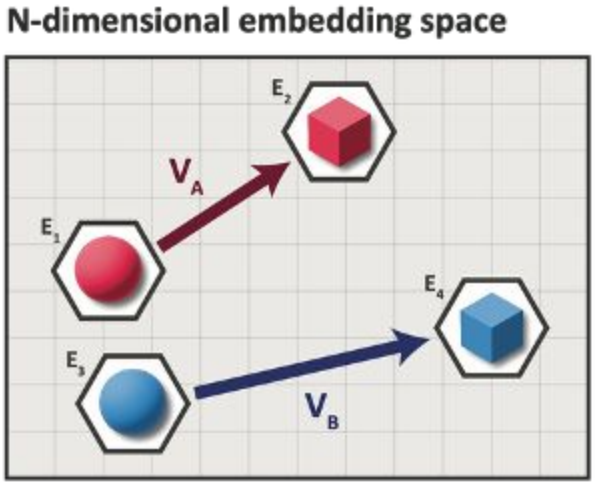

Lower NMI consistently correlated with better generalization.

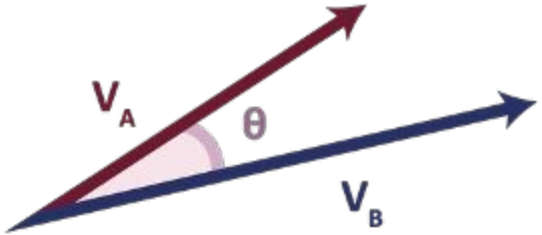
# Why Does Diversity Improve Generalization?




But, for the same NMI, more diverse data still led to better generalization.

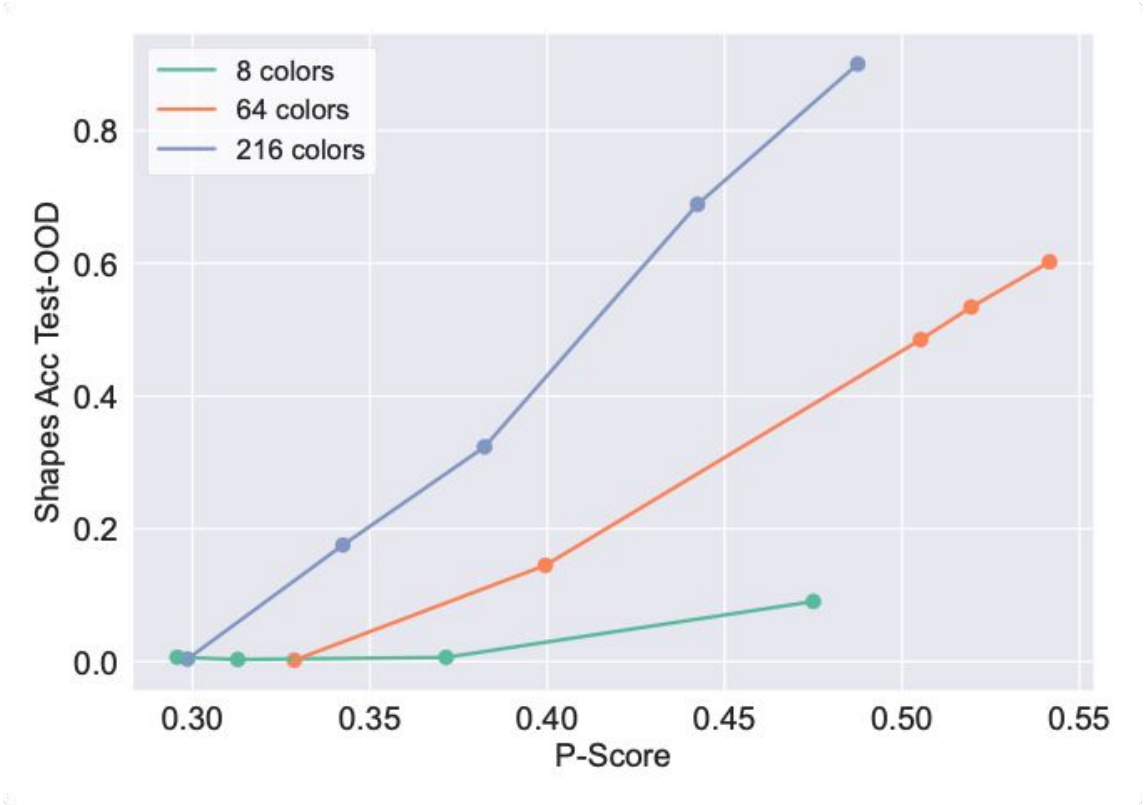
# Exploring Relationship between Diversity and Parallel Representations



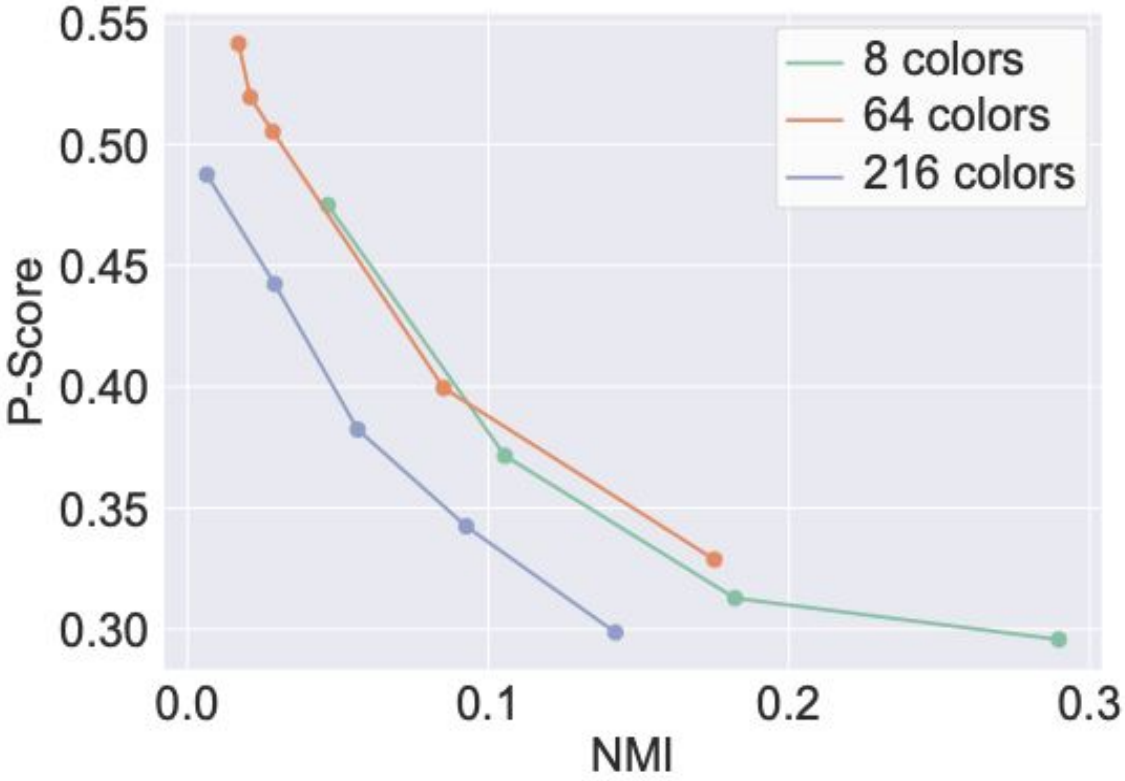




$$p\_score = cos(\theta) = \frac{V_A \cdot V_B}{||V_A||\ ||V_B||}$$

# Parallelism Links to Systematic Generalization



# Parallelism Links to Systematic Generalization



# Takeaways

- Diversity, Burstiness and Latent Intervention promotes SG in a Multimodal Masked Language Modeling task.

- Diversity improves SG independently of reduced NMI.

- NMI promotes SG by promoting the emergence of representations with more parallelism under attribute changes.