Beyond next-token prediction

Joint Embedding Predictive Architecture (JEPA)
for Cardiac Vision-Language Understanding

Apply via TUM Thesis Management Portal

Transformer models

PREDICTING THE NEXT TOKEN

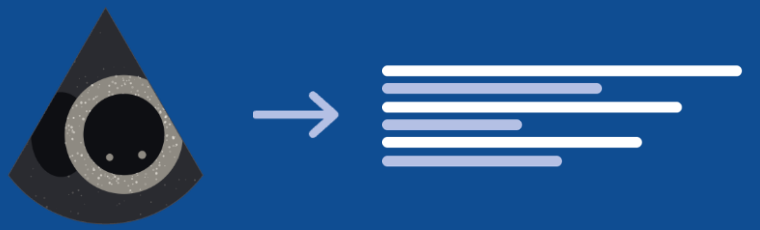
LVEF	is	55	%	healthy
LVEF	is	25	%	severe heart failure

LVEF = share of blood pumped per beat. “55” and “25” are one token apart, so the loss barely notices.

Grades words, not meaning

JEPA-VL (this MSc Thesis)

PREDICTING THE MEANING



Instead of words, it predicts an embedding of the heartbeat, so a healthy heart and a failing heart land far apart.

Ignores speckle noise

Learns from context

Marta Hasny · Sameer Ambekar
Cristina González · Prof. Dr. Julia Schnabel

CompAI Lab at TU Munich