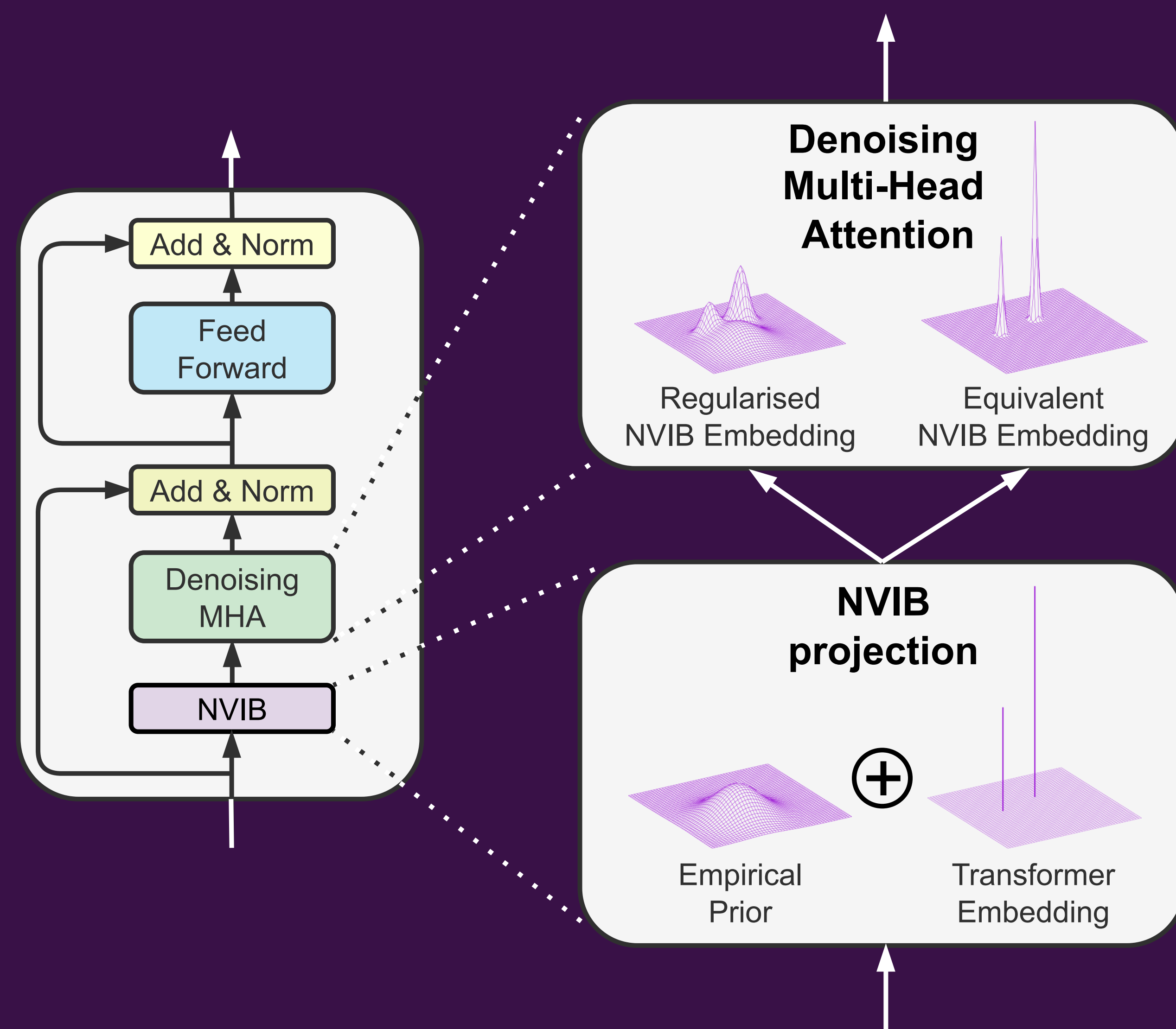
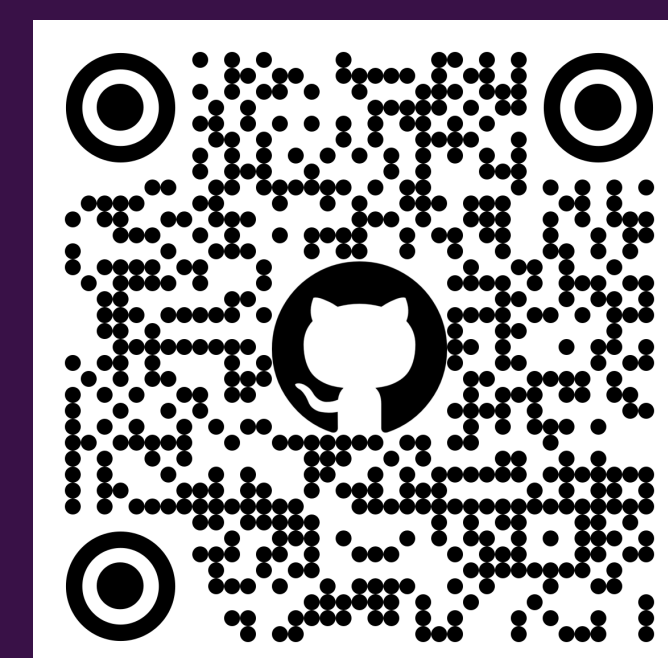


Use information theory to  
reinterpret **Transformers**  
and improve performance  
**without training!**



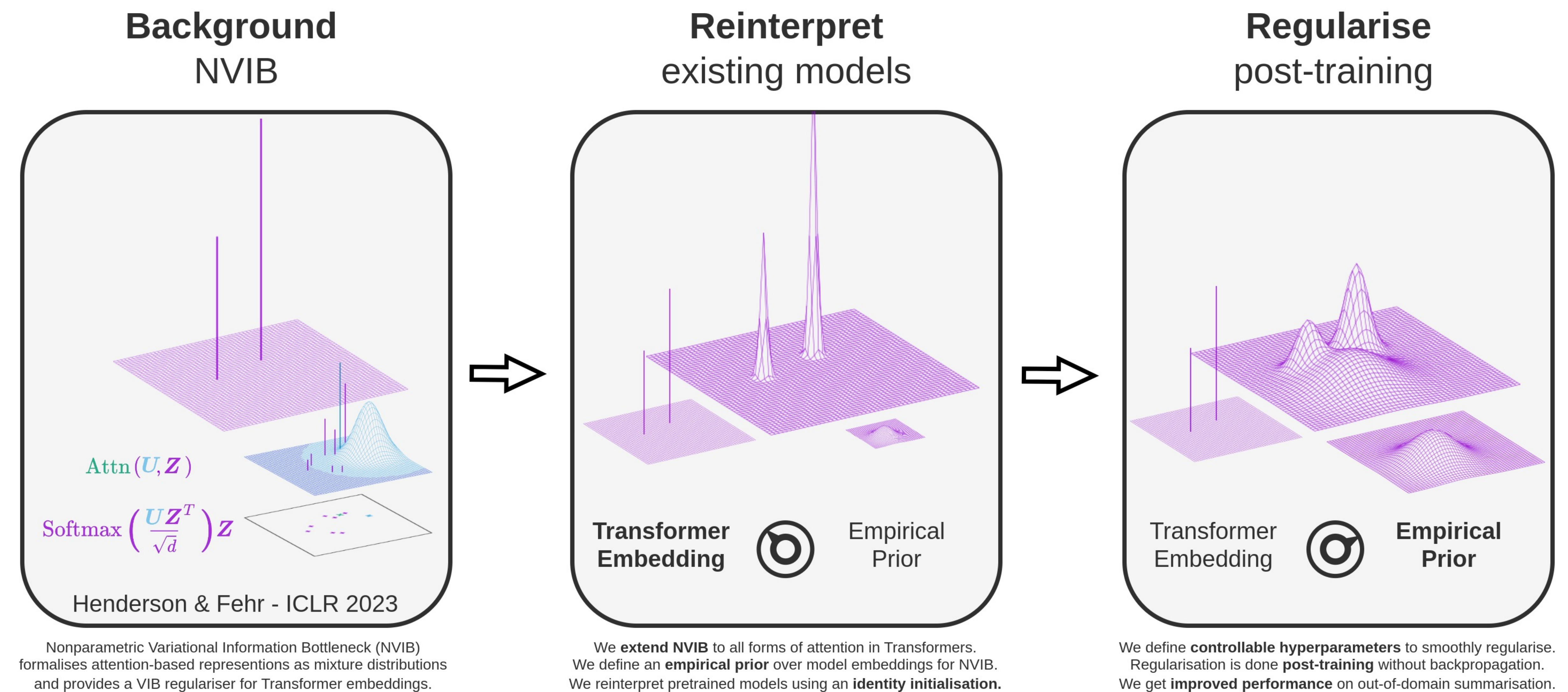
Paper



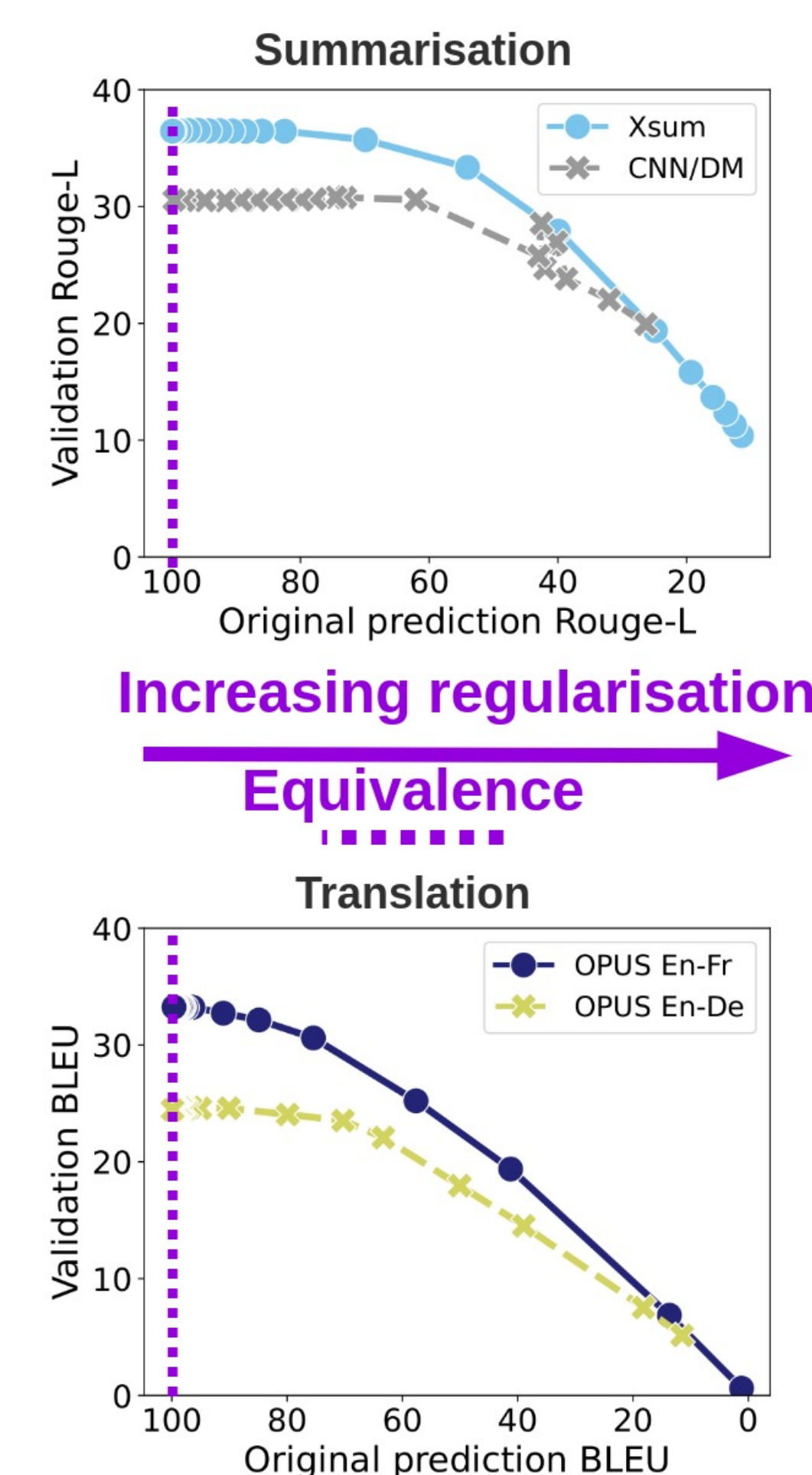
Code

# Nonparametric Variational Regularisation of Pretrained Transformers

Fabio Fehr & James Henderson



## Smooth regularisation



## Improved out-of-domain generalization

| Summarisation               |         | ↑% |
|-----------------------------|---------|----|
| NV-BART <sub>(Xsum)</sub>   | +18.2 % |    |
| NV-BART <sub>(CNN/DM)</sub> | +2.3 %  |    |
| Translation                 |         | ↑% |
| NV-Marian <sub>En-Fr</sub>  | +1.0 %  |    |
| NV-Marian <sub>En-De</sub>  | +0.4 %  |    |

Relative improvement over baseline

## Adaptive generation

