Logit lens
eval aux loss