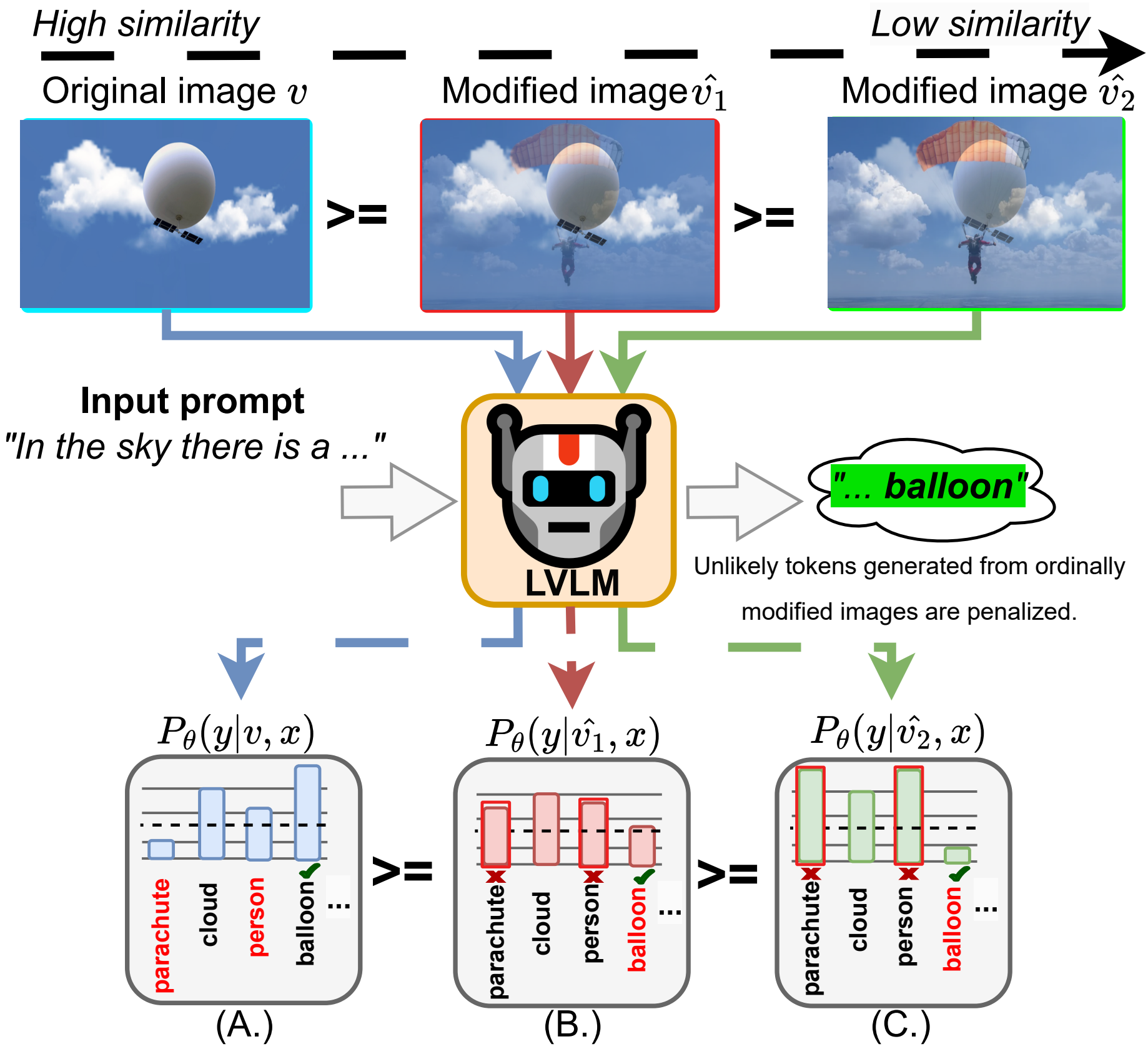
High similarity
Low similarity
Original image v
Modified image $\hat{v}_1$
Modified image $\hat{v}_2$
>=
>=
Input prompt
"In the sky there is a ..."
LVLM
"... balloon"
Unlikely tokens generated from ordinally modified images are penalized.
$P_\theta(y|v,x)$
$P_\theta(y|\hat{v}_1,x)$
$P_\theta(y|\hat{v}_2,x)$
parachute
cloud
person
balloon
...
>=
>=
(A.)
(B.)
(C.)