Inputs Happy Father's Day **Self-Attention**  $\overline{o}_{self}$ **Cached Attention**  $\overline{o}_{mem}$