- For instance: each number is treated as a separate token
- Replace them with a number token $\textsc{num}$
  - but: we want our language model to prefer

$$p_{\text{LM}}(\text{I pay 950.00 in May 2007}) > p_{\text{LM}}(\text{I pay 2007 in May 950.00})$$

  - not possible with number token

$$p_{\text{LM}}(\text{I pay } \textsc{num} \text{ in May } \textsc{num}) = p_{\text{LM}}(\text{I pay } \textsc{num} \text{ in May } \textsc{num})$$

- Replace each digit (with unique symbol, e.g., $@$ or 5), retain some distinctions

$$p_{\text{LM}}(\text{I pay 555.55 in May 5555}) > p_{\text{LM}}(\text{I pay 5555 in May 555.55})$$