LSTM
or long short-term memory cell

Let's compare LSTM and RNN

RNN

Zooming In

\[ S_t, W_y, \bar{S}_{t+1}, \bar{y}_{t+1}, W_x, W_s, \bar{y}_{t+1}, \bar{x}_{t+1}, W_h \]
Zooming In

\[ x_t \]

\[ S_t \]

\[ \sigma(x) \geq 1 \quad \text{All data passes through} \]

\[ \sigma(x) < 0 \quad \text{No data passes through} \]

Gating functions

They decide what data to retain.
LSTM Explained (visually!)

Reminder for RNN

$$M_t = \tanh \left( W [STM_{t-1}, E_t] + b \right)$$
So now let's see how LSTM looks like:

Let's look at each component individually.
\[ f_t = \sigma (W_f [STM_{t-1}, E_t] + b_f) \]

\[ LTM_t = LTM_{t-1} \cdot f_t + N \cdot i_t \]
\[
U_t = \tanh \left( W_u LTM_{t-1} \cdot f_t + b_u \right)
\]

\[
V_t = \sigma \left( W_u \left[ STM_{t-1}, E_t \right] + b_v \right)
\]

\[
STM_t = U_t \cdot V_t
\]

\[
(\text{output}_t)
\]