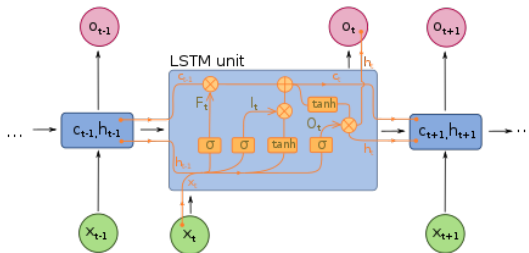


# Deep Learning par la Pratique

Long-Short Term Memory

---

# Long-Short Term Memory



- $F_t = \sigma(W_F * x_t + U_F * h_{t-1} + b_F)$  (forget gate)
- $I_t = \sigma(W_I * x_t + U_I * h_{t-1} + b_I)$  (input gate)
- $O_t = \sigma(W_O * x_t + U_O * h_{t-1} + b_O)$  (output gate)
- $c_t = F_t \odot c_{t-1} + I_t \odot \tanh(W_c * x_t + U_c * h_{t-1} + b_c)$
- $h_t = O_t \odot \tanh(c_t)$
- $o_t = f(W_o * h_t + b_o)$

# Long-Short Term Memory

vanishing gradient “résolu” (ou presque)

