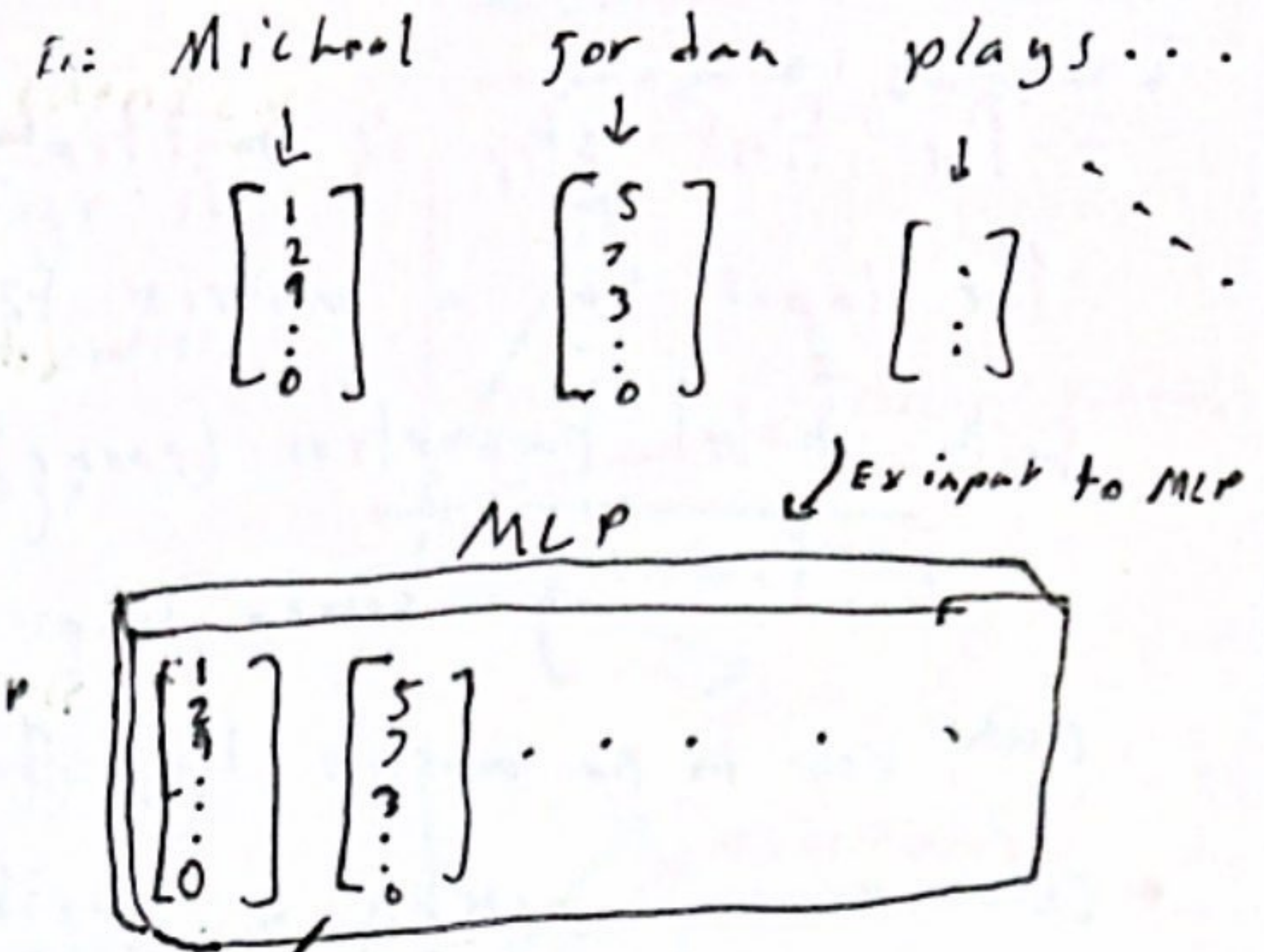
MLP (P-3) (LLM notes)

- inside a MLP: the sequence of vectors which where each vector is associated with a input token.



- Then each vector in that sequence inside the MLP goes throung a sequence of operations

Ex for one vector $\to \text{linear} \begin{bmatrix}2\\5\\9\\\vdots\\0\end{bmatrix} \to \text{RELU} \begin{bmatrix}3\\5\\9\\\vdots\\0\end{bmatrix} \to \text{linear} \begin{bmatrix}0\\1\\7\\\vdots\\2\end{bmatrix}$

- at the end we get a new updated vector this gets added to the original input vector in the MLP and that sum is the output that flows out

$$\begin{bmatrix}0\\1\\7\\\vdots\\2\end{bmatrix} + \begin{bmatrix}5\\7\\3\\\vdots\\0\end{bmatrix} = \begin{bmatrix}5\\8\\10\\\vdots\\2\end{bmatrix} \to \text{output}$$

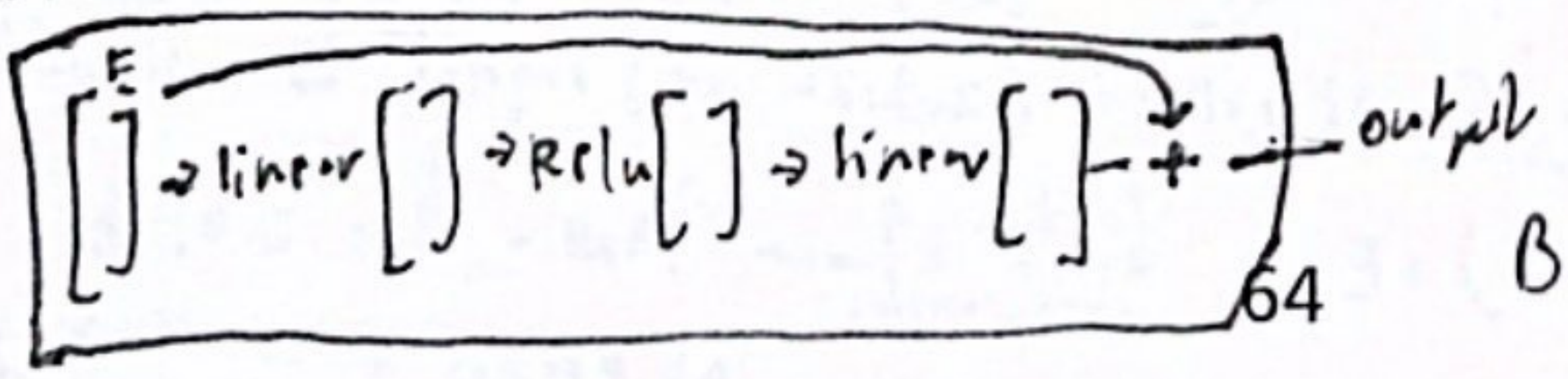
- you then apply this operation to all vectors in the sequence ie for every token in the input, all this is in parrallel and indipendent of one another

- What this means is that if the input "Jordan" includes context of first name michael last name Jordan then the last operation before adding will produce a vector for the direction "Basket Ball" which will be added to the vector for Jordan in the input, this produces the result MJ + basket Ball which gives the token Jordan context of being associated with basket ball

Michael Jordan → Note vector Jordan has context of MJ from attention block (see last page)

$$\begin{bmatrix}5\\7\\3\\\vdots\\0\end{bmatrix} \xrightarrow{\text{linear}} \begin{bmatrix}\vdots\end{bmatrix} \xrightarrow{\text{Relu}} \begin{bmatrix}\vdots\end{bmatrix} \xrightarrow{\text{linear}} \underbrace{\begin{bmatrix}0\\1\\7\\\vdots\\2\end{bmatrix}}_{\text{BasketBall}} + \underbrace{\begin{bmatrix}5\\7\\3\\\vdots\\0\end{bmatrix}}_{\text{MJ}} = \text{output} \begin{bmatrix}5\\8\\10\\\vdots\\2\end{bmatrix}$$

MLP:



E
linear
Relu
linear
+
output