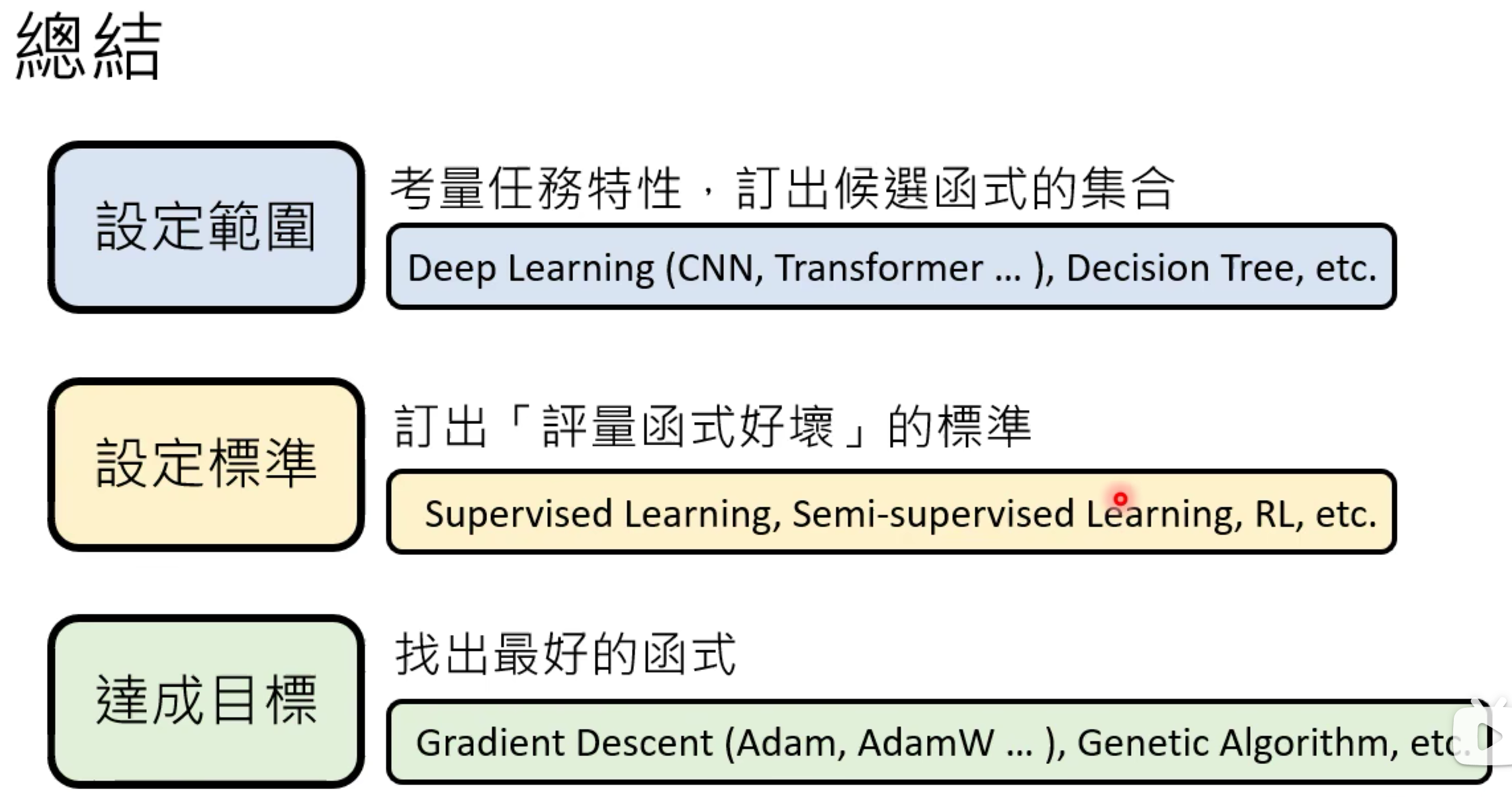
## 李宏毅2023春机器学习课程

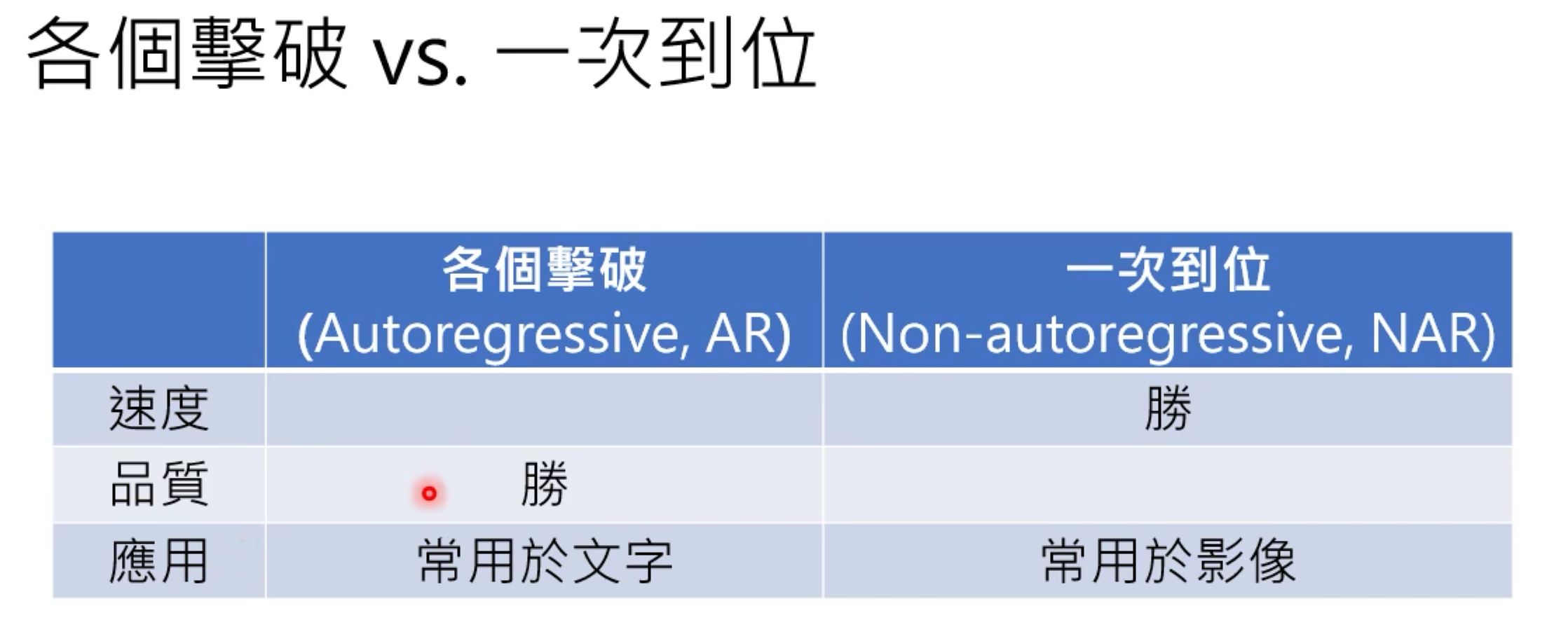
P14



其中，划定的函数集合大小与训练数据量是有关系的。更大的训练数据量可以采用更大的函数集合范围，就如近年常有采用自注意力效果优于CNN的例子，其中一个原因就是目前CV方面数据量已经比较大，合适采用有更大的函数集合的自注意力方法。

P15

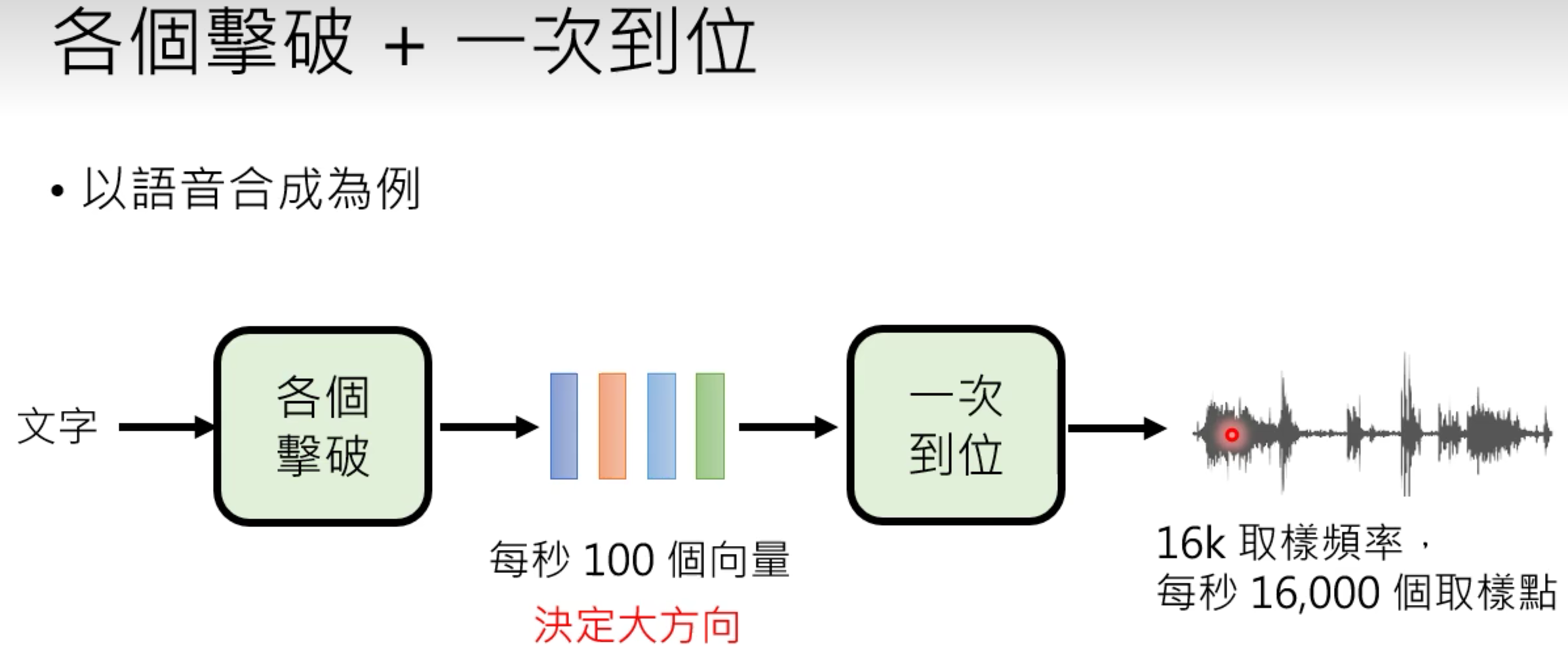
①生成式模型的两种策略：



前者如chatgpt，生成文字时一个一个生成，相对质量更好，但速度更慢。

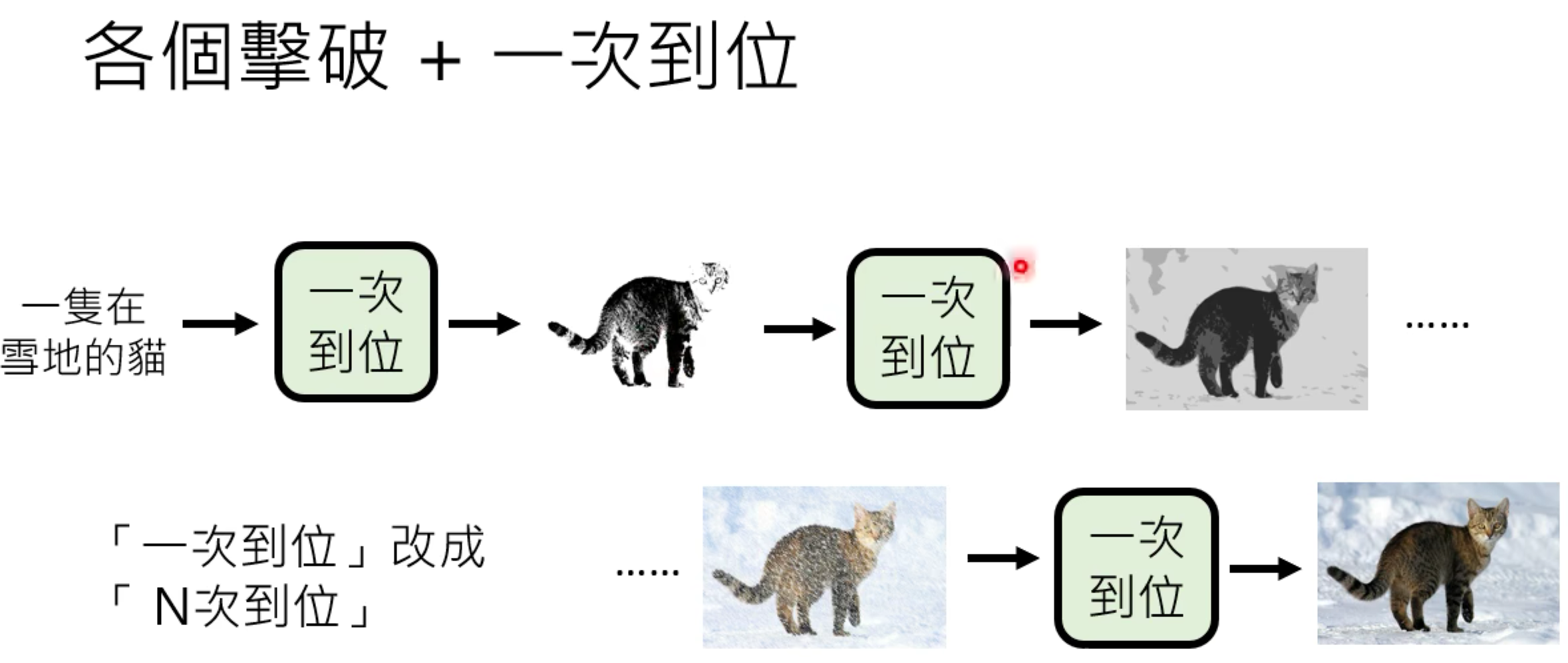
因此可以把两种方法结合起来使用：

①



先用逐个击破的策略生成较为简略的中间产物，决定大方向，再一次到位生成结果。

②

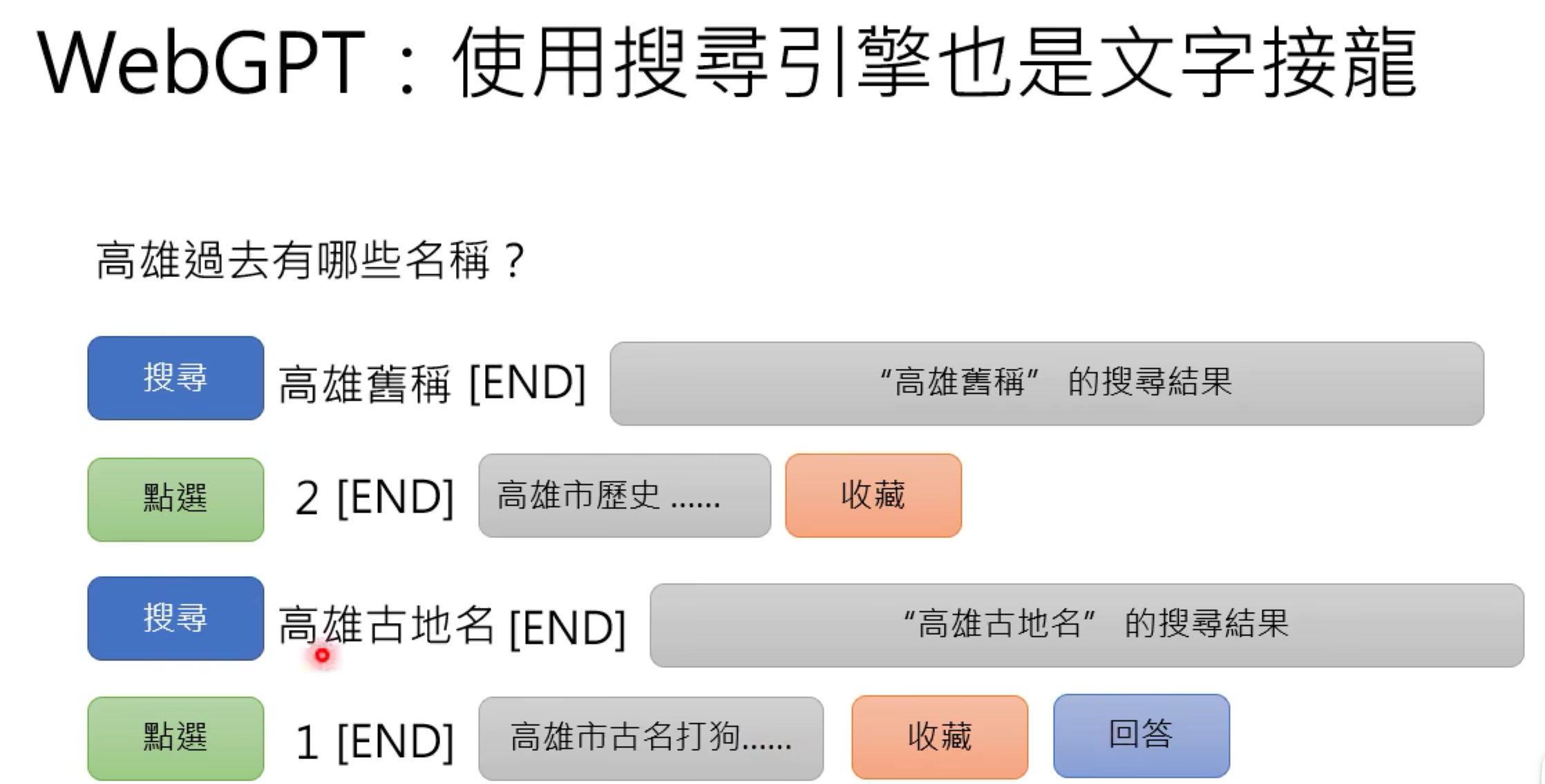


采用N次生成的方法，也结合了两种策略。（diffusion model）

P16

①New Bing：与Chatgpt最大的不同在于可以搜索网络

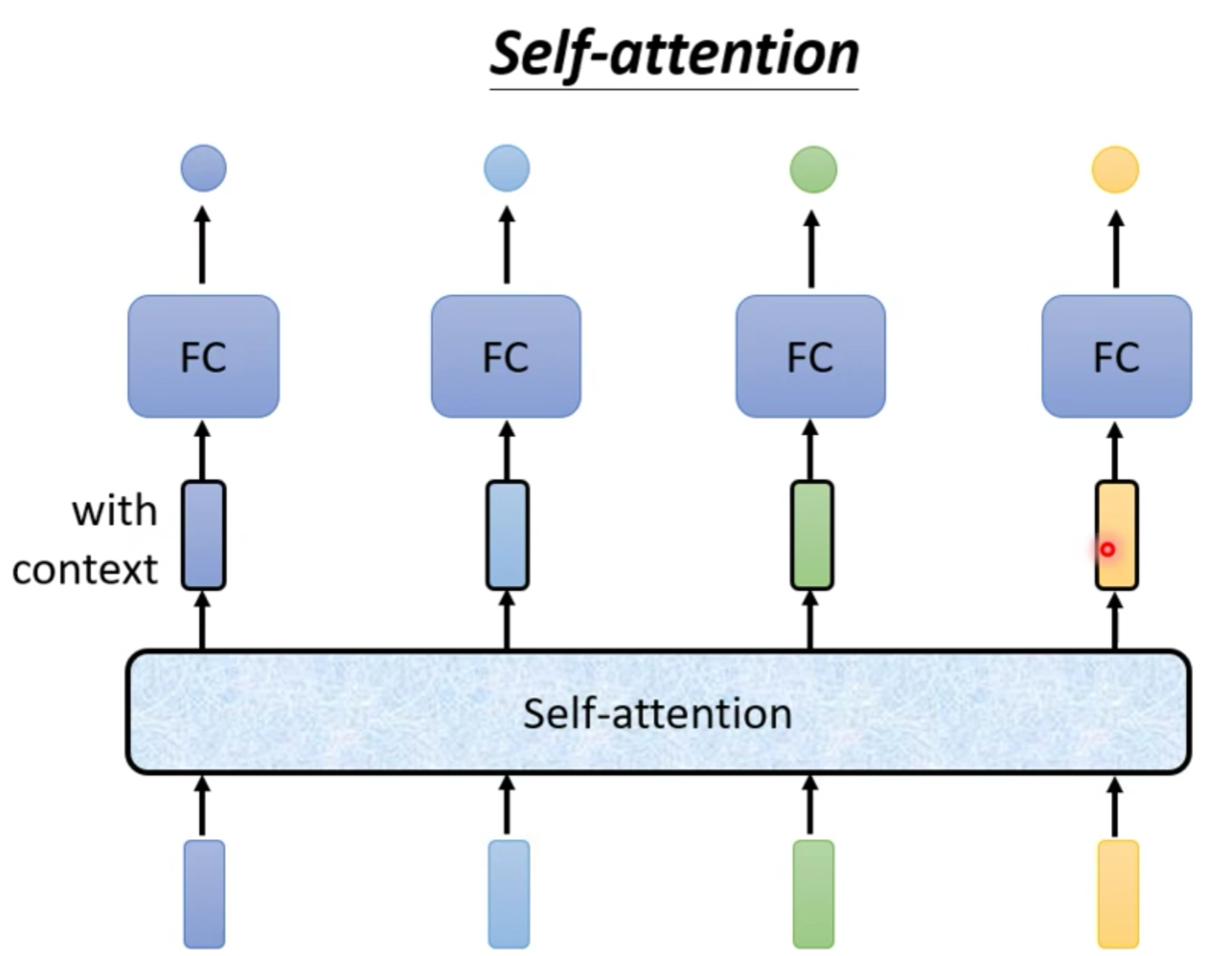
②WebGPT与New Bing相似。使用搜索引擎也是在做文字接龙，其大致流程如下所示：



③Toolformer：除了使用搜索引擎，还可以使用多种不同的工具，如计算机，翻译等。

P21

自注意力机制的大致思路：每一个输入序列中的元素会与同一序列中的其他所有元素进行交互。通过这些交互，模型能够根据每个元素的相关性来调整其表示。



对于序列中的一个向量，计算其他向量与其的相似度（关联程度）。根据关联分数抽取重要信息。大致步骤：

**①Query, Key 和 Value 向量（Q, K, V）**:

Query（查询）: 用来检索相关信息的向量。

Key（键）: 用来匹配查询的向量。

Value（值）: 实际存储信息的向量。

对于每一个输入元素，都会计算出一个Query向量，一个Key向量和一个Value向量。

**②注意力分数**:

每个Query向量与所有Key向量计算相似度（点积），得到一个权重分数。

这些分数被用来表示输入序列中不同元素之间的相对重要性。

**③加权求和**:

使用Softmax函数将注意力分数转换成概率分布。

使用这些概率对相应的Value向量进行加权求和，得到最终的输出表示。

其中，计算Q, K, V的过程中，需要的矩阵W是模型要训练的参数。

Linux命令测试：

