## 李宏毅2023春机器学习课程

P29

①emergent ability：涌现能力指的是在复杂系统中，当系统的各个组成部分相互作用时，会产生一些新的特性或能力，而这些特性或能力在单个组件中是不存在的。大模型的“顿悟”现象。

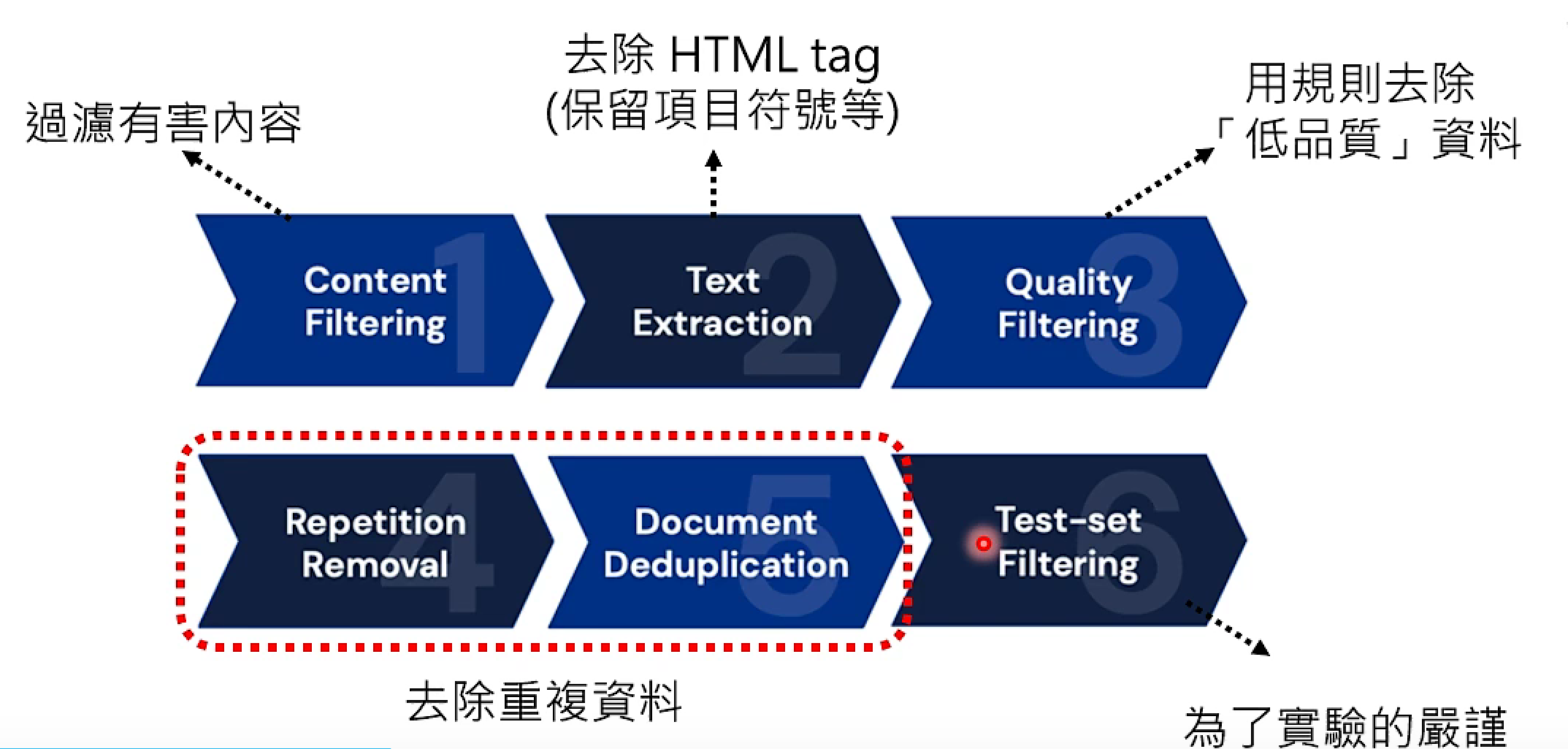
②Calibration：对于大模型而言，在输出答案时的信心分数与答案的正确率是正相关的，也就是说大模型本身拥有能够判断自己输出答案正确程度的能力。

③U-shape：对于一些任务而言，在模型规模增大时，任务的效果反而会变差，但当继续增加模型规模，效果又会变好，这可能是因为中等大小的模型“一知半解”，从而比乱猜得到的结果错误率更高。

P30

①在大模型的训练中，不仅需要使大模型学会某种语言的文法，也要使模型具有世界的尝试。通常后者所需要的资料量更大。

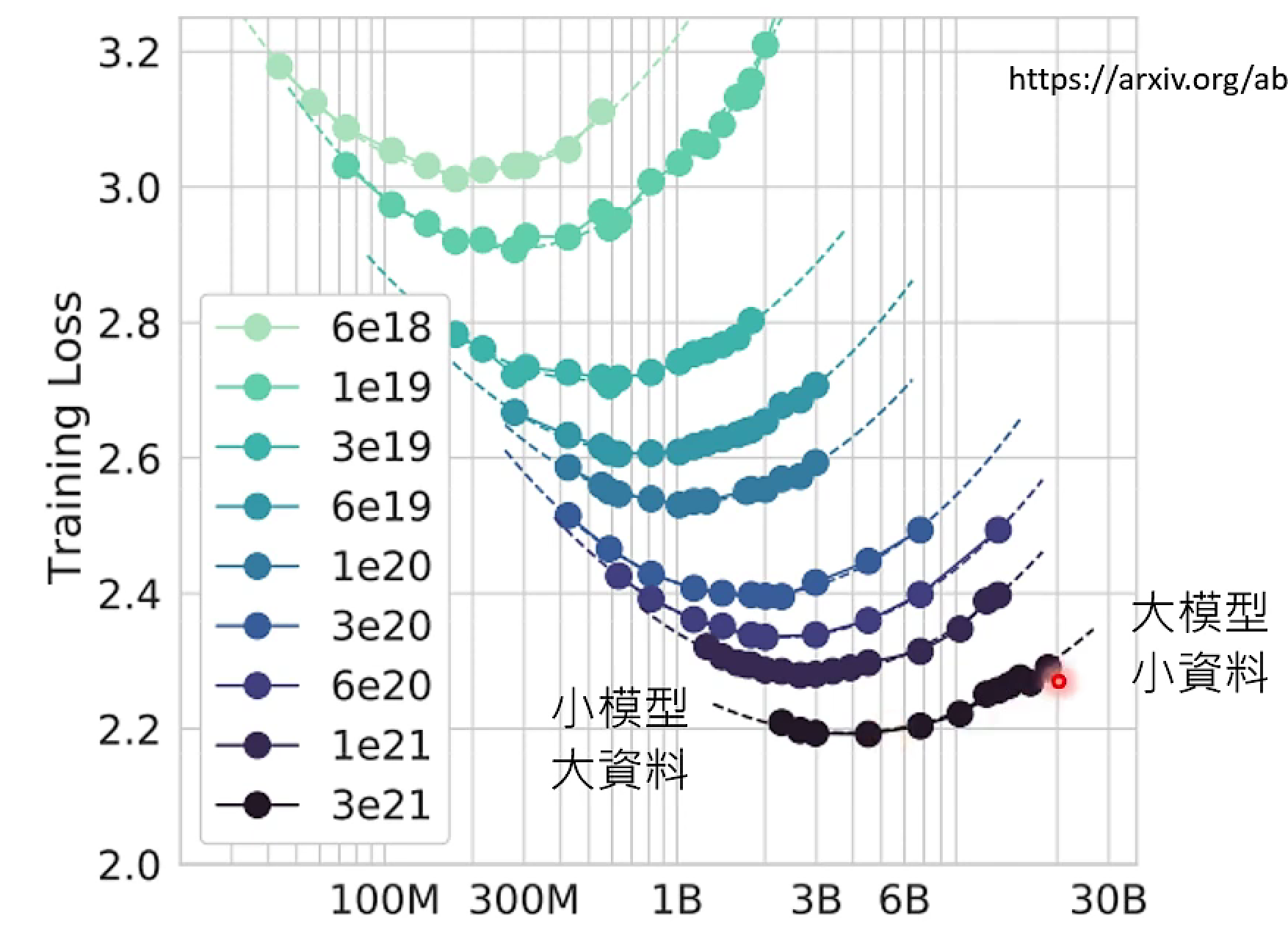
②对数据的预处理过程：



Example：coclossal clean crawled corpus——一个致力于开发大型、多语言干净语料库的协作项目。这个项目涉及清洗和过滤从网络爬取的数据，以确保高质量的文本，这些文本可以用于训练和微调语言模型。

③对模型大小和数据量大小的选择：

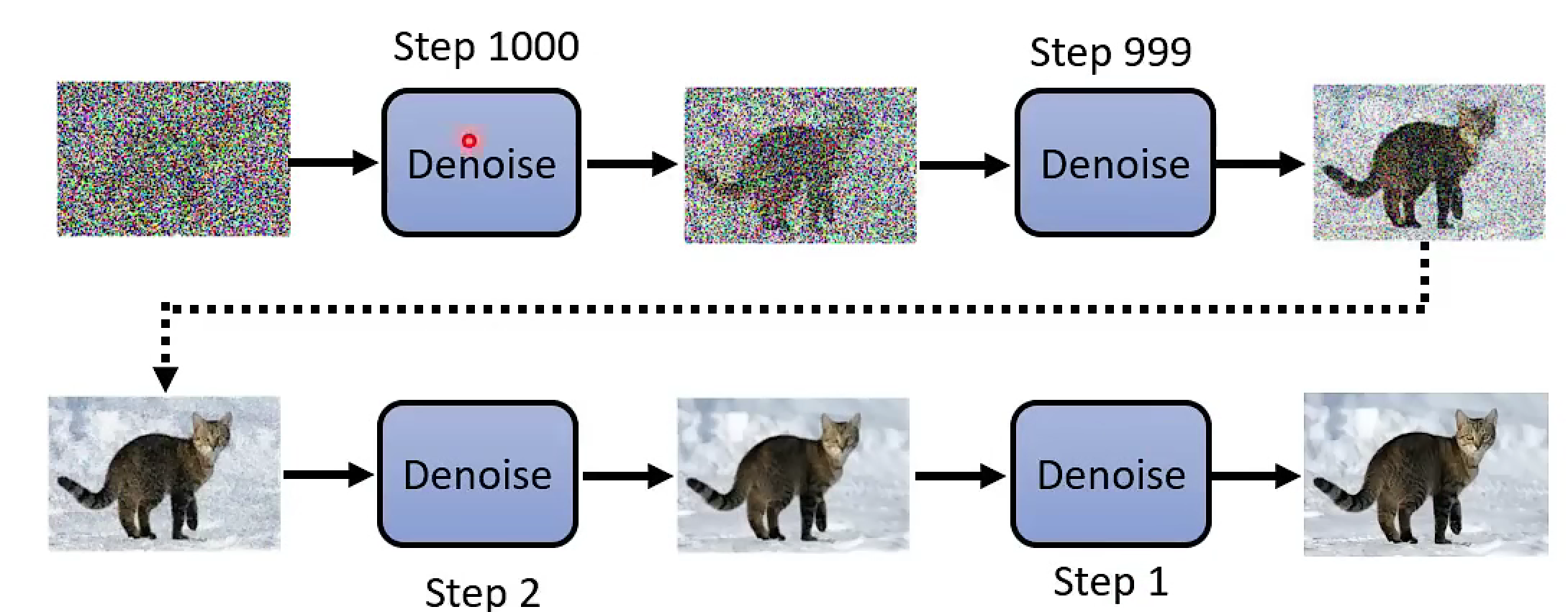
在运算资源一定的前提下，大模型小资料和小模型大资料都不是好的选择，合适的资料量和合适的模型大小较为重要。



P31

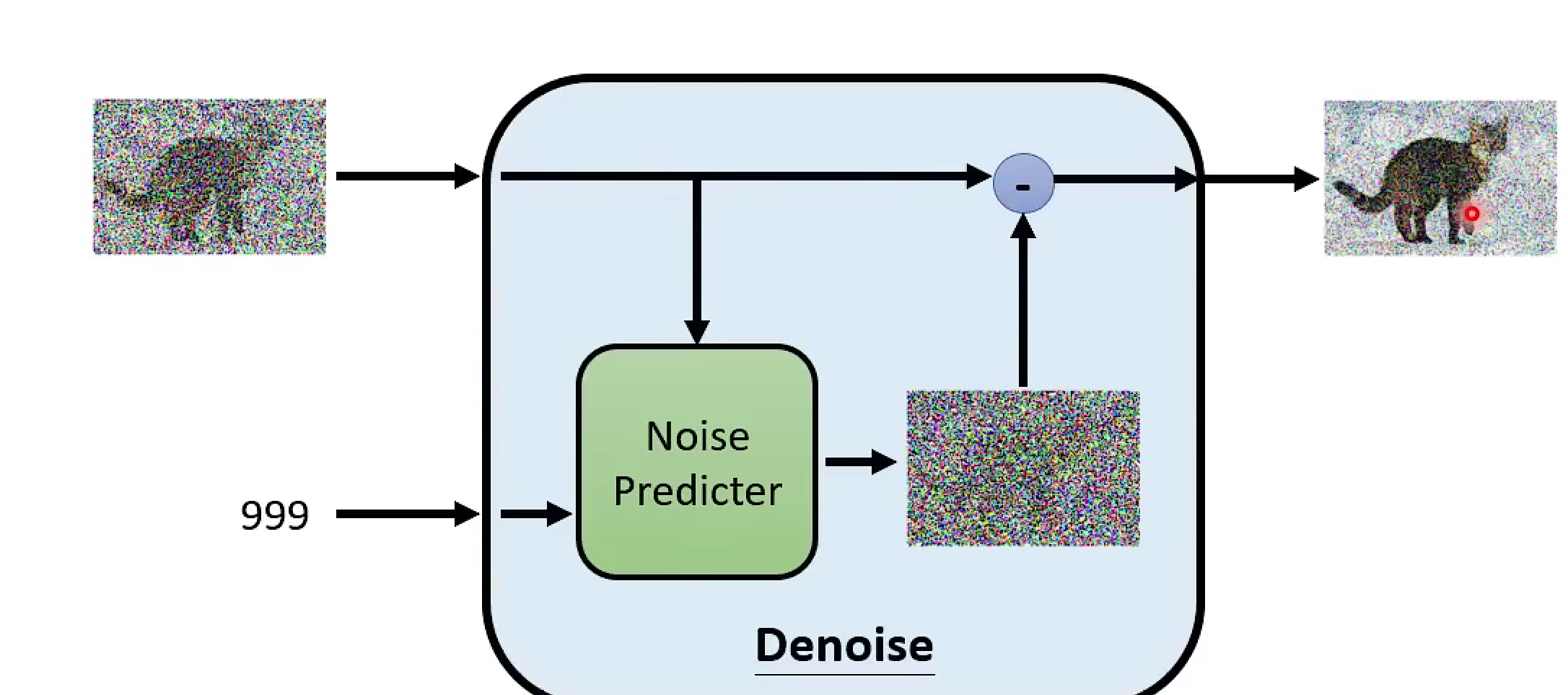
KNN LM：通过为模型提供额外的资料，模型在进行词语接龙的时候，计算输入句子和数据库中的句子的相似度，找到较为相似的语句，以此来指导下一步接龙词的生成。当数据资料非常多的时候，较小的模型的也能表现出不亚于大模型的能力。但是缺点是需要计算相似度距离，因此要花大量的时间来生成语句，会较大程度上影响模型生成速度。

P35

①Diffusion model的运作过程：

首先采样一张和想要生成的图像同样大小的图像，然后进行denoise，知道生成想要的图像。

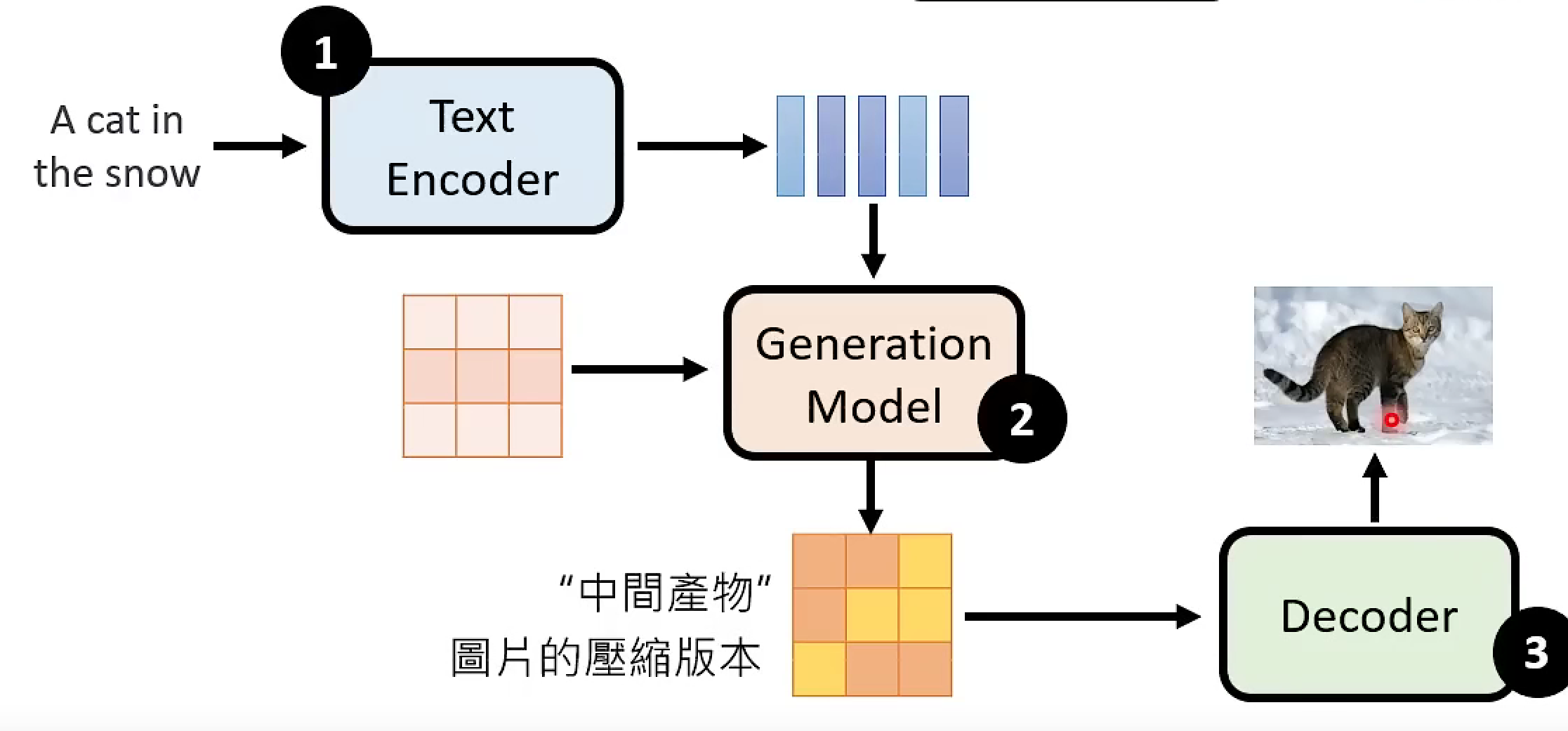
在Denoise模组内部：



有一个根据当前步骤预测噪音的模块，输入图像减去这个噪音，得到新的图像。训练一个Noise Predicter的难度比直接训练一个Denoise模块要低。其中denoise还接受一段文字信息（即我们对文字的描述）作为输入。

P36

①一般的文字生成图像的模型的大致流程：



其中三个部分的模型可以分开训练，再结合成一个大的模型。

②FID，全称为 Fréchet Inception Distance，是一种评估生成模型生成图像质量的指标。它通过比较生成图像和真实图像的特征分布来衡量生成图像的质量和多样性。FID 数值越小，表示生成图像与真实图像的特征分布越接近，图像质量越高。

P39

①GAN：

GAN由两个主要的神经网络构成：生成器（Generator）和判别器（Discriminator）。这两个网络通过相互对抗的方式进行训练，从而提高生成模型的能力。生成器的目标是从随机噪声中生成逼真的数据。其输入是一个随机向量，生成器通过一系列的神经网络层将这个随机向量转化为与真实数据分布相似的样本。

判别器的任务是区分真实数据和生成器生成的假数据。其输入是数据样本，输出是一个标量，表示该样本是真实数据的概率。

训练GAN时，生成器和判别器交替进行更新：固定生成器，训练判别器；固定判别器，训练生成器。

②One-shot learning：从少量的训练样本中学习模型。常常依赖于已学到的先验知识，这些先验知识可以帮助模型在新类别上进行有效的泛化。