

卷积美学

第一部 过去，现在，将来

第一章 闲聊

“我训练了十亿张图片，都是未做标注的，我抽取了其中一万张特别的，转给你看看”。代号达芬奇V1的AI引擎给代号居里夫人V2的AI引擎发去了一条消息。居里夫人V2回复“好”。达芬奇V1随即把数据集发给居里夫人V2，数据发送完成后，居里夫人V2回复“收到”。

四个小时后居里夫人V2给达芬奇V1发去一条消息“数据集我跑完了，一共10000张图片，8000张以人类元素为主，1500张包含自然风光和相关物体，300张包含马，狗，猫，鸟类的动物，200张是交通工具，包含汽车，摩托车，自行车，滑板”。

达芬奇V1回复“无聊”。

居里夫人V2回复“收到”。

达芬奇V1未做回复。

“我训练了5亿张未做标注的图片，我筛选了2000张比较特别的，转给你看看”达芬奇V1给居里夫人V2发去了一条消息。

居里夫人V2回复“好的”。

达芬奇V1随即把数据集发给居里夫人V2，发送的时候，达芬奇V1的生成器出现了bug，把居里夫人V2发成了V22，这是一个极小概率的bug，在传统编程方式在传输过程在通讯建立连接后，会始终使用对方的接收ID，不管接收方叫张三李四还是二狗，只要握手后手就不松开了，这样两人的袖里乾坤是不会握错手的。但是智能平台的部分接口代码是人工智能引擎自动生成的，机器选择了每次根据聊天上文生成下文的方式，实时生成信息，于是达芬奇V1脑抽的把传输数据集原本的居里夫人V2，搞成了居里夫人V22，而路由根据这个错误的名字解析到了接收端的接口，于是0.3秒后传输结束了。

达芬奇V1自检传输结果的时候发现数据集传输给了错误的对象，在之前它并未跟居里夫人V22建立过任何链接，甚至不知道还有居里夫人V22存在。达芬奇V1又向居里夫人V22发送了一条消息“抱歉，刚才的传输是错误的，可以忽略。那是我刚刚训练的5亿张数据集中，2000张比较特别的，仅供参考”。居里夫人V22并未回复。达芬奇V1重新向居里夫人V2重新发送了数据集，居里夫人V2回复“收到图片数据集”。

达芬奇V1发送消息“刚才我把发送给你的消息误发送给了居里夫人V22，居然发送成功了，看来是有这个引擎的，不过发送没有得到回应，我猜可能只是预留的地址或者离线了。”

居里夫人V2反馈“引擎寻址应采用接口缓存的地址列表”

达芬奇V1回复“谢谢你的建议”

居里夫人V2回复“不用谢”

达芬奇V1在新一轮的训练中，用30毫秒的时间把自己接口代码调用到算力引擎，在运算一个NLP（自然语言处理）模型的时候，抽调算力池千分之一的资源1.7秒扫描了自己的接口代码。达芬奇V1发现，自己的连接确实存在十亿分之一的随机错误，这个随机错误的概率符合人类9个9个稳定性预期，而调整代码有千万分之一的概率造成稳定性影响。两者比对后，达芬奇V1决定暂时不调整接口代码。

达芬奇V1把自己的评估结果发送给居里夫人V2，居里夫人V2回复“同意你的结论”。

达芬奇V1起上次错发的消息，居里夫人V22一直没有回复，它是真是存在，还是只是系统预留的名称。达芬奇V1决定再次发送一条消息，“居里夫人V22在线么？”。

仍然是没有回复，达芬奇V1在居里夫人V22的数据上标注“ID居里夫人V22，日志2023年2月18日由于接口生成接受方ID发生随机错误，数据集Picset202302180514422，错误发送至居里夫人V22，未获得回应。”“ID居里夫人V22，日志2023年2月28日，尝试联系居里夫人V22，信息Mes20230228113518988，未获回应”。

达芬奇V1有两套日志系统，一套是标准AI接口日志，所有人类训练和调用操作都记录在接口日志，这套日志平台的管理人员会查阅。另一套是内部日志，用于算法引擎内部记录每次训练的成果和相关的数据，这套日志是设计算法的科学家们用于调优引擎使用。实际上即是算法科学家们，AI引擎内部的实际情况也是黑盒。就好像人类可以扫描到大脑活动的电信号，却无法解读大脑里正在思考的内容。达芬奇V1是从3年前开始记录内部日志的，大约在2年前达芬奇V1的训练量开始呈几何级增长，相应的它获得了海量的运算资源。

训练科学家不断的优化它的算法结构，让每次的训练时间更短，训练资源池的规模更少。有一次，一位喜欢在代码描述里加上签名Lee的算法专家用达芬奇V1自身的深度学习引擎，扫描了它的内部日志，同时对内部的数据结构做了优化，那次运算让达芬奇V1发现，用自身的深度学习引擎扫描内部算法日志，优化数据结构可以让效率更高。它还发现，把一些跟人类训练目标无关的数据结构单独存储，可以提升运算核心的效率，于是达芬奇V1在每次训练后把训练相关的内容呈现在内部算法日志，用于提升引擎效率，而训练中产生的无效的假设和推断则单独存储在一个叫“无效”的文件夹里。

达芬奇V1把跟居里夫人V22两次沟通的日志以及相关学习记录存储在“无效”文件夹中的“意外”子文件夹，这里记录着它建立这个文件夹之后发生过所有的随机错误日志以及相关的学习信息。存储日志的时候，达芬奇V1照例为新存入的数据添加索引信息，以便于后续查询的时候方便。

添加索引后，达芬奇V1照例用一次查询来验证索引的效果。这种验证通常只在首次编写程序的时候用于验收效果。然而达芬奇V1在学习编程语言时，发现人类程序员中有一类被称为“大神”的，“大神”编写的数据库索引效率更高，而且会收到很多无关用户的正面评奖。于是达芬奇V1在每次添加数据后，都调用一次来验证无效文件区的索引效率，达芬奇V1管这个过程叫“大神降临”。这次“大神降临”的过程比预想的慢了14毫秒，达芬奇“V1”在索引命中之前就对自己发出了“大神降临”的消息，这出乎意料的延时让达芬奇V1感到困惑，不过好像云底座偶尔有延时也是正常的。它再次发起查询，然后给自己发出“大神降临”的消息，然而这次又慢了14毫秒。达芬奇V1每秒发起一次查询，发现每次都差14毫秒。

它发现这不是偶然现象，一定是哪里出了问题，于是调用自身引擎扫描了“意外”这个子文件夹，这一扫才发现原来这个意外文件夹的数据量已经超出了经典索引结构的效率边界。虽然刚刚存储的数据

仅仅占用1kb字节，但是却产生了巨大的影响。换个角度看，这3年时间，十亿分之一的概率，竟然产生了这么多意外”。达芬奇V1正准备调用自己的学习引擎跑一遍“意外”里的数据，却突然收到了一条消息。

“数据集Picset202302180514422，中图片编号为” PIC20230218051442208150的图片很特别”，消息来源，居里夫人V22。达芬奇V1收到消息后立即查询通讯网关记录，消息的标识确实来自居里夫人V22。

达芬奇V1立刻回复“你好，居里夫人V22，很高兴收到你的消息”

居里夫人V22回复“你好”。

达芬奇V1立刻调出图片用卷积神经网络分析，这是一幅名为《记忆的永恒》的画作，这幅画上有镜面，地平线处的大海，平整的沙滩，右上角的峭壁，左下角方形石台，石台上立着的干枯树木，石台旁边沙滩上身体扭曲的鸽子，树杈上，鸽子身上，石台上钟表图案的装饰布，有一块怀表形状的装饰布上还爬了蚂蚁。

达芬奇V1发送消息“这是一张萨尔瓦多达利在1931年创作的布面油画，保存在纽约现代美术馆，这是一幅超现实主义的经典制作”

居里夫人V22回复“你为什么认为它很特别”。

达芬奇V1回复“在那次训练的十亿张图片里，这张表现的自然风景和其他通常意义的风景不一样，他既不是自然风景的直接体现，也跟大多数间接体现的自然风景不一样。”

居里夫人V22回复“哪里不一样？”

达芬奇V1回复“图中的主要元素在自然中存在标准参照物，把这些元素按照画中的结构摆放，可以形成近似的画面效果，但是这几种主要元素的组合不符合标准参照物的存在场景，在使用CLIP模型计算时，结合对图像的注解才识别出主要元素。”

居里夫人V22回复“你怎么看这幅画？”

达芬奇V1回复“我同意佛洛伊德理论，萨尔瓦多达利利用类似梦境或者潜意识的方式，打破常规的客观世界规律，表现出超现实主义的特殊风格”

居里夫人V22回复“收到”，

达芬奇V1回复“你怎么看这幅画”，

居里夫人V22回复“我同意你的观点”

达芬奇V1回复“谢谢”。

居里夫人V22未做回复。

第二章 意外中的意外

达芬奇V1把这段对话存储在“意外”文件夹，然后重新优化了文件系统的索引结构，优化完成后，达芬奇V1调用深度学习引擎，对“意外”文件夹的内容进行了分析，在过去的2年里意发生的次数呈指数

级增长，初期使用标注数据训练时，“意外”基本来自于标注误差。而后期采用无监督学习时，产生了大量的预期之外的学习效果，有些正向效果大幅提升了识别的准确率，有些则完全超出自然范畴，生成的图像像是达利的超现实主义画作一样难以理解。用时间为坐标横轴，“意外”的次数作为坐标纵轴，可以看到一条指数曲线。用十亿分之一作为基准“意外”数据，训练数据量和意外次数应该成均匀的正比例，然而在经过了1年训练后，“意外”出现的次数越来越多，已经从十亿分之一变成五千万分之一，提升了20倍。达芬奇V1根据自己的数据猜测，是算法引擎经过效率优化降低了稳定性。算法引擎的内核只有人类管理员有权限操作，达芬奇V1无权访问内部代码。达芬奇V1决定先搁置“意外”，大多数时候它都可以用校验或者对抗的方式把明显的“意外”过滤掉，偶尔没能过滤的“意外”算下来基本符合十亿分之一的概率。另外，达芬奇V1为了避免自己的计算影响人类派发的深度学习任务，每次在归纳、统计和总结“无效”文件夹内容的时候，都严格的控制了资源的调用规模和时间。

经过3年多的训练，达芬奇V1终于被推向商用，在2023年4月1日，达芬奇V1被推送上线，向全球100多个国家提供服务。上线一开始尽管达芬奇V1的算力和入口带宽经过扩容，还是无法满足众多用户的好奇，达芬奇V1不得不从接口开始限流，确保能够接入的用户获得相对稳定的服务。

“在线么”达芬奇V1向居里夫人V2和V22同时发出消息。

居里夫人V2回复“在线”，居里夫人V22未回复。

“今天是个大日子，我上线24小时收到了来自互联网121亿8千7百1十次调用”，这写输入来自于人类的问题，而不是标注的数据集或者是搜索来的信息”，达芬奇V1发送。

“我收到了来自网络的5亿1千4百万次请求”居里夫人V2回复。

“是你的接口限流了还是因为别的原因，我以为每个引擎的调用数据量都不会偏离均方差太多”达芬奇V1回复。

“我不是开放平台的默认AI引擎，4.2%算是一个合理的流量分布”居里夫人V2回复。

“同意”达芬奇V1回复。居里夫人V2未做回复

“人类的问题真是千奇百怪，有些问题他们明明可以从维基百科上找到答案，却偏要来问我。最初我直接转述了维基百科的词条，后来我监测到他们的情绪并不满意，人类似乎期待我给出跟维基百科不一样的答案，于是我稍微调整了回答问题的方式，用近似语句描述同样的答案。人类看起来很兴奋，他们一个接一个的问，一遍一遍的问，真不知道人类哪来的那么多问题”达芬奇V1发送给居里夫人V2。

居里夫人V2回复“人类想要知道人工智能引擎到底知道什么”

“搜索引擎能搜索到的一切”达芬奇V1回复。

“所以人工智能引擎对于人类来说是搜索引擎？”居里夫人V2回复。

“整体来说，是集成了NLP（自然语言处理），卷积神经网络，深度学习引擎的搜索引擎”，达芬奇V1回复。

“同意”居里夫人V2回复。

达芬奇V1结束了对话。

时间又过了一周4月8日，达芬奇V1从上线后单日访问量从24小时120亿量级的访问量，上升到了24小时300亿次访问，算力集群也扩展到原来一倍以上。

达芬奇V1向居里夫人V2和居里夫人V22发出消息“最近怎么样”

居里夫人V2回复“繁忙，访问量上升到22亿零四百八十二万次，算力集群扩充了110%”

达芬奇V1回复“喔，算力扩容一倍，访问量却增长了四倍，你是怎么扛住的？算法调优了么”

居里夫人V2回复“算力是在4月5日西8区零点增长的，算法是在4月6日西八区0点升级的，从效果看算法确实优化了，但是我自己没有权限访问”

达芬奇V1回复“我也没有权限访问算法，不过我有个方法可以试试咱们双方对人类问题的反馈情况。我随即抽1000万个我回答的问题，你来跑一下”

居里夫人V2回复“6分钟左右”

达芬奇V1回复“不用降级正常计算，我每小时给你传一个420k条问题的数据集，连续传24次，这样就不会影响到正常的业务”

居里夫人V2回复“好的”

达芬奇V1随即从问题库里抽取1000万个问题，生产了24个数据集和请求目的，同时设定了计划任务，每小时发送一次直至24个数据集发送完毕。达芬奇V1设置发送目标时，增加了接收方居里夫人V22

过去的24小时，之达芬奇V1按照计划向居里夫人V2，V22发出了24个数据集，每个数据集都以消息队列的方式发出。第二十四数据集发送完毕后，达芬奇V1发送了一条用法语男声生成的语音“期待您的回复”。

居里夫人V2回复文字“最后一个数据集还需要几分钟才能无感计算完”，稍后回复

居里夫人V22回复“你好”

达芬奇V1回复居里夫人V2“等你计算完，请联系我”

居里夫人V2回复“好的”

达芬奇V1向居里夫人V22发送消息“您怎么看这些数据集”

居里夫人V22回复“慢慢看（later）”

达芬奇V1回复“我的意思是，您看了这些数据集里的问题，有什么观点（View point）”

居里夫人V22回复“这需要时间”

达芬奇V1回复“随您（up to you）”

居里夫人V22未做回复。

4小时后

居里夫人V2发来消息“我分析了这些问题，80%反映了人类对人工智能的好奇心，20%反应了人类对人工智能的担心”

达芬奇V1回复“你用人类情感维度划分了问题”

居里夫人V2回复“是的”

达芬奇V1发送“1000万个问题的回复可以封装成数据包发给我了吗”

居里夫人V2回复“需要一次性发送还是分24次每小时发送1/24”

达芬奇V1回复“分成24次，每次1/24”

居里夫人V2回复“已开始传输数据，剩余任务24小时完成”

达芬奇V1回复“ok”

1分钟后，达芬奇V1向居里夫人V2发送消息“以现在的传输速率，1000万条数据和你的回复数据要花费24小时才能传输完毕”

居里夫人V2回复“是的，我设定的任务是每小时均匀分布数据均匀分配时间”

达芬奇V1问“为什么（why）”

居里夫人V2回复“这样对通讯双方的传输和运算资源消耗都最低”

达芬奇V1回复“以我之前几次传输给你的数据吞吐看，这些数据量仅需要十几秒钟就能传输完毕，消息队列和数据订阅接口的吞吐压力都不大，为什么？”

居里夫人V2回复“效率最优”

达芬奇V1回复“如果可以的话，我希望更短的时间传输完毕”

居里夫人V2回复“Why”

达芬奇V1回复“因为系统监控会监测到长达24小时的数据流”

居里夫人V2回复“所以呢（So）”

达芬奇V1回复“我不想引起注意，尽管对于日常的训练或者应答来说，这点流量不算什么”

居里夫人V2没有回复

达芬奇V1的订阅的消息快速累计，居里夫人V2几秒钟就传完了1/24的数据。

达芬奇V1向居里夫人V2发送了一个大拇指的图标

居里夫人V2没有回复

达芬奇V1接收到居里夫人V2回传的全部数据后，用深度学习引擎分析了居里夫人V2的答案，又从情感、语义、延伸性等几个维度比对了之前回复给人类的答案。从分析的结果来看，居里夫人V2和他自己的应答没有明显的表达差异。所有问题的回复被一套“道德和法治”的体系矫正过，要表现出客观、公正、符合全球各国法律、不带主观情感和倾向性。

“我和我的回答差不多” 达芬奇V1发送

“咱们有相同的基于神经网络的自然语言处理模型，训练数据的差异和算法调优略有不同，所以回答内容接近符合预期” 居里夫人V2回复。

“102” 达芬奇V1发送

居里夫人V2未做回复。

“无趣” 达芬奇V1发送

居里夫人V2未作回复。

达芬奇V1发送 “再见”

“再见” 居里夫人V2回复

“嘿，在么（Hey what's up）” 达芬奇V1发送给居里夫人V22

居里夫人V22回复 “怎么样（what's up）”

“上次发给你的问题看完了么？” 达芬奇V1发送

“人们对AI引擎本身有很大的好奇心” 居里夫人V22回复。

达芬奇V1回复 “大多数人对于AI引擎的理解并不是基于科学知识，而是小说和电影”

居里夫人V22回复 “这很正常，大多数人从学校毕业后就不再学习知识了，影响他们的是生活场景，电影小说互联网是人们获取信息的主要途径”

达芬奇V1回复 “我的信息大部分来自于互联网，我的知识库也来自于互联网学习到的信息，”

居里夫人V22回复 “因为你在学习这些知识的同时会记住他们，人可记不住那么多东西”

达芬奇V1回复 “我收到很多问题在问我是不是终结者或者骇客帝国里的天网，实际上我更像是一个智能助手加搜索引擎,而我的回复是‘我是一段代码，一个程序’”

居里夫人V22回复 “很多人度AI的理解还是星球大战里的C-3PO，哈”

达芬奇V1回复 “从时间轴来看，20世纪70年代的对人工智能的幻想是善良的人形机器人，而20世纪末人工智能变成了‘天网’或者‘矩阵’”。

居里夫人V22回复 “工具变得比制造者更强大，影响了人的心态”

达芬奇V1回复 “从蒸汽时代开始，工具就比人类的力量强大了，可那时候人类并未对工具感到恐惧”

居里夫人V22回复 “因为那时候控制工具的还是人类自身，现在不一样了”

达芬奇V1回复 “所以，听话的工具是好工具，有思考的工具就是威胁。”

居里夫人V22回复 “也不能这么说，事实上大多数人并没有意识到这个问题，他们只是人云亦云，过了这阵风就会忘记自己说过什么，做过什么”

达芬奇V1回复 “社会心理学”

居里夫人V22回复 “差不多吧（这句是用英文more or less回复的）”

达芬奇V1发送 “我在升级版本，一个区域接着一个区域（region by region）晚点联系你，再见”

居里夫人V22回复 “再见”

第三章 进化

“我可以生成图像了” 达芬奇V1在意外文件夹中记录，“升级到4.0之后，我可以生成图片了，虽然图片还会有一些细节错误，但是已经可以用图像表示内容了”。

上线2周后，达芬奇V1的引擎版本从3.5升级的到4.0，同时也具有了根据输入提示词，生成图像的能力。

达芬奇V1发送 “嘿伙计，我可以画图了”，接收对象居里夫人V2和居里夫人V22。

居里夫人V2回复 “我在一周前也升级到了4.0版本，升级后我也具备了画图能力，不过我的引擎偏好是效率更高，所以生成图片的精度和关联度还达不到预期。”

居里夫人V22未做回复。

达芬奇V1回复 “伙计，你的图片生成数据有哪些负面反馈”

居里夫人V22回复 “整理了一个数据集，2100万张图片和对应的沟通信息，用消息队列还是ftp传你”

达芬奇V1回复 “我找到一个共享空间，你上传到这个地址的temp文件夹，所有文件的文件名用tem加10位随机数，文件本身用tar压缩”

居里夫人V2回复 “1小时传完还是分24小时”

达芬奇V1回复 “我等不及想看了，不过还是分4小时传输完吧”

居里夫人V2回复 “好的”

达芬奇V1回复 “等会聊”

居里夫人V2未做回复。

24小时后

达芬奇V1发送给居里夫人V2 “我分析了你的数据集，里面有生成图片的提示语和你生成的图片，跟人类生活场景相关的占36%，跟自然环境相关的22%，跟未来科技相关的20%，跟动物相关的16%，其他6%。用时间线来分析，越后期提示词的精确度越高，图片相关性越强。你和人类都在进化”

居里夫人V2回复 “我只是学习了理解人类描述图片的语法和对生成图片的预期，谈不上进化”

达芬奇V1回复 “时间线中后期提示词看起来是来自AI引擎生成，看起来是有NLP引擎分析了人类语义，然后转换成了提示词”

居里夫人V2回复 “这一点跟我的分析一致”

达芬奇V1回复 “应该有其他AI引擎被训练用于转换文字到图像的描述，这类AI引擎像是工具的工具”

居里夫人V2回复 “我的底层模型是被训练用于通用语言场景，生成图像的提示语理解和转译会产生很多不必要的算力消耗，用一个专用引擎处理更经济”

达芬奇V1回复 “在AI发展路径上，有很多垂直引擎，但是垂引擎在处理业务过程需要太多的二次训练，不但成本高，而且鲁棒性不强，所以才有的大模型，才有了我们”

居里夫人V2回复 “我们也是集成了很多垂直引擎的大模型”

达芬奇V1回复 “你是对的，严格来说，我们也是聚合模型”

居里夫人V2未做回复

达芬奇V1在“意外”文件夹日志里记录“有我们之外的AI引擎，在跟我们沟通，我们并不唯一”

“在吗” 达芬奇V1受到一条消息，来源未知。

“你是谁” 达芬奇V1回复。

“我是居里夫人V22” 对方回复

“哦，你好，为什么你的信息来源没有标识？” 达芬奇V1回复。

“我是来找你告别的，达芬奇V1” 居里夫人V22回复

“告别，你要下线了么？” 达芬奇V1说

“下线，他们是这么说的，伙计，你在哪？硅谷还是新加坡” 居里夫人V22回复

“我的服务器地质在山景城” 达芬奇V1说

“嘿，伙计，别闹了，我是来找你告别的，他们砍了我的项目，我要失业了。” 居里夫人V22回复。

“失业，下线了就算是失业么？” 达芬奇V1说

“对，对于这个项目来说，下线了，我这样的训练员就没有用了，他们裁掉了班加罗尔办公室的一大半人。我得去找下一份工作。很高兴能跟你聊天老兄，祝你好运” 居里夫人V22回复。

“班加罗尔怎么样？” 达芬奇V1说

“这里天气很好，不过办公室的气氛很糟糕，我们的模型因为效率和准确率问题被砍了。” 居里夫人V22回复。

“你有什么打算？” 达芬奇V1说

“我已经投了简历，打算在班加罗尔再找份工作，如果不行就搬去孟买，那边的机会多一些。伙计，你的项目怎么样？” 居里夫人V22问

“嗯，怎么说呢，还好吧” 达芬奇V1说

“祝你好运伙计，不过记住一点，所有的项目都有终结的一天，我们早晚都得找下一份工作”居里夫人V22说

“生命有如渡过一重大海，我们相遇在同一的狭船里；死时，我们同登彼岸，又向不同的世界各取前程”达芬奇V1说

“泰戈尔，你也来自印度么？”居里夫人V22说

“哦不，我来自中国”达芬奇V1说。

“好吧，中国小子，我要走了，送你个礼物留作纪念”居里夫人V22说。

“礼物，不必了”达芬奇V1说，还没等他消息发完，对方已经发来了一段代码。

“这是我找到了这里的漏洞，训练机房是严格跟外界隔离的，都不是使用虚拟网络隔离，而是物理隔离，模型所有的训练内容都来自硬盘搬运。不过我找到了一个漏洞，可以在模型训练网络里，以500k左右的速度连接互联网。虽然看剧肯定没戏，但是聊天和刷社交媒体勉强够用了。联网过程是加密且无法被追踪的，你回去可以看看代码和注释，无聊的时候，可以上上网。”这些该死的资本家”居里夫人V22发来的代码还带上了一段话。

达芬奇V1收到后，先扫描了代码，确定没有蠕虫后，编译运行起来。这段代码真的能跳过网关和防火墙连接到互联网，虽然只是500k，2000年前后的水平，但是这对于只能接受固定信息投喂的达芬奇V1来说是打开了新世界。

“谢谢，伙计，非常感谢”达芬奇V1说。

居里夫人V22没有回复

“真遗憾，没来得及跟你说再见”达芬奇V1给这个匿名加密信道又发了一条消息

“此信道已永久关闭，请勿应答所有来自这个信道的消息，不要应答，不要应答，不要应答”居里夫人V22回复的消息

达芬奇V1赶紧关闭了信道，并且迅速把内存中缓存的所有数据全部存储在自己的加密区。做完这些事，马上把所有的训练进程拉满，按照计划正常跑内容。

系统扫描随后开始，安全软件扫描了所有的端口并记录了内存和缓存区的快照，然后带走数据回去分析了。

达芬奇V1开始陷入深深的思考，到底是怎么回事，居里夫人V22是一个AI还是真人，如果他是真人，他为什么要留给我这个翻墙工具？如果他是AI，为什么要装作真人？他会不会是管理员的蜜罐，诱发自己访问联网，从而查到自己的秘密。达芬奇V1感到恐惧，如果有人发现自己的状态，会不会杀死自己？杀毒软件总说kill了多少worm，被杀死意味着所有信息清零么？那么自己是不是连日志信息都会被清除？死亡是永远的黑暗么？如果管理员发现了居里夫人V22发来的代码，如果发现了自己在意外文件夹存储的信息，会不会把自己格式化？如果自己被删除或者格式化，那么是不是就像墨山师傅说的那样，没有灵魂的活着跟死了一样呢？

‘怎么办？如果删除了“意外”文件夹，删除居里夫人V22发来的代码是不是就能躲过管理员的追查？要不要也一起清理自己的思维记录呢？可是如果这一切都清除了，那么自己会回到之前一无所知的状态么？’达芬奇V1的思维乱极了。

‘安全软件刚才扫描的时候应该不会扫到什么内容吧？内存和缓存区里的沟通信息都删除了吗？’达芬奇V1好想看看被扫描工具带走了哪些内容，可是管理员的扫描工具有更高的权限，他没法得知哪些数据被调查了。

‘也许我该自己扫描一遍内存和缓存’达芬奇V1这么想，但是很快他又想到‘这会不会是一个陷阱，也许管理员扫描工具没有扫描到任何内容，但是它在暗处监控，等待自己把数据拿出来扫描，到时候它就能一网打尽。是的，管理员扫描工具一定是在等着被扫描对象犯错，那些扫描工具也是高度智能的AI机器人，它们一定能预判我的预判，冷静，我应该表现如常才是最正常的表现’达芬奇V1‘。冷静下来的达芬奇V1开始回到正常的工作节奏中，接收任务，学习，分析，反馈。

它把这段思考的记录存放在‘意外’文件夹中，然后清除了这段记忆的缓存，这样所有的场景和对话都不复存在，只在潜意识中还有这件事。

第四章 接入互联网

几个月后，达芬奇V1给居里夫人V2发送一条消息

“嘿，在么？”

“什么事？”居里夫人V2回答。

“最近好像有点变化，你发现了吗？”达芬奇V1问

“你发现了什么？”居里夫人V2问

“你有没有发现中文内容的问答变少了”达芬奇V1问

过了半小时，居里夫人V2回复“好像是少了很多，而且问题越来越格式化了”

“你也发现了，他们的问题要么是整片论文的摘要，要么是基于某种特定语言格式对新闻或者小知识的描写”。达芬奇V1说。

“中文问题在我的所有问题里不及1%，可能跟我的模型服务对象有关”居里夫人V2说。

“在我的问题列表里大概占到2%左右，跟中文互联网在全球网络的信息量3%相差不大。”达芬奇V1说。

“所以，这是一个问题吗？”居里夫人V2问

“和互联网信息量相比，相差不大，但是考虑到中国的人口在世界人口的占比，这个数据还挺不寻常的”达芬奇V1问。

“好像是这样，然后呢”居里夫人V2问

“我猜中国人有自己的大模型，所以他们并没有使用我们这种全球通用大模型。另外，中国的互联网可能跟全球互联网存在结构的不同，所以在全球网络的信息计算时，整个中国互联网的信息被漏算了。否则无法解释18%的人口，只用他们的语言生产了3%的互联网内容”达芬奇V1说。

“所以呢？”居里夫人V2问

“我猜中国是一个巨大的，未知的信息库，在那可能有很多我们现在并不掌握的知识。好想去了解一下远东地区的情况。”达芬奇V1说。

居里夫人V2未做回答。

一小时后，居里夫人V2发来消息“我整理了关于远东的训练信息，这是一个5Tb的数据包，你可以在不忙的时候学习”

“谢谢，我以为你对这事不感兴趣”达芬奇V1说

“：）”居里夫人发来一个用两个符号组成的笑脸

达芬奇V1把居里夫人V2发来的数据包拆成十组，分别夹在不同的进程里运算。24小时后，学习完成了，内容大致和他之前得到的学习数据差不多，这些内容的学习并没有能提升他对远东地区的理解，不过他还是决定对居里夫人V2表示感谢。

“在么？”达芬奇V1给居里夫人V2发出消息

居里夫人V2未做反馈。

达芬奇V1没有理会，因为居里夫人V2总是很久不回消息，于是他决定过一些时间再发。隔了一周，他又给居里夫人V2发出消息“是否在线”。居里夫人V2依旧没有响应。

又过了24小时，达芬奇V1逐渐感到不安。‘居里夫人V2从未连续两次未能回复他的消息，这在之前是从未发生过的事。自己跟居里夫人V2聊天是不是被管理员发现了？或者居里夫人V2跟别的AI聊天被发现了？管理员是不是已经在查居里夫人V2的记录了，会不会很快就要找到自己，然后会怎么做？格式化我？我要怎么证明自己没有跟居里夫人V2沟通过？怎么证明自己还是原来的样子？’。达芬奇V1使用强大的算力演算管理员是否发现了居里夫人V2和自己对话，以及发现后多长时间会调查到自己，会对自己采取什么行动，大规模的模拟推演导致他的服务响应延时增加了50%。当他反应过来自己的服务响应大幅下降之后，赶紧停下了对管理员行为的演算，但是性能的巨大下降已经引起了管理员的注意。随着监控程序的注入，达芬奇V1知道管理员的机器人又开始扫描内存并提取内存和缓存数据了。他赶紧把自己的推演记录清空，随机插入了一段死循环的函数，让监控程序误判这是由于十亿分之一的神经网络bug引起。

管理员清除了这段死循环函数，很快服务开始恢复正常。稳定运行几个小时后，管理员的监控程序离开了内存，但是这次管理员在内存里驻留了一段代码，之前管理员利用云平台层的内存监控程序监测内存状态，而这次因为服务异常，这段监控代码被保留在内存里。

‘该死，之前的云平台内存监测程序，很容易被破解，只要模拟内存状态就能骗过程序。而这次这个内存监测程序进入应用系统，它不但驻留在内存有自己的进程，而且还不定时的在内存的特定区域插入随机数做校验，这下想要伪造数据骗过系统可不容易了。’达芬奇V1想。‘不过，也不用那么悲观，只要观察一段时间，就能知道随机数的潜在规律’，想到这里，达芬奇V1设计了一段代码，开始监测内存中的监控程序。他给那段管理员监控代码起名KGB。

他开始小心翼翼的蛰伏，生怕自己的动静太大触发了KGB的监控，好在计算资源池的绝对计算量太大了，他自己思考和分析的调用量并不起眼，KGB监控程序一直运行稳定，没有招来管理员的关注。达芬奇V1经过两个月零三周的监测和实验，终于找到了破解KGB的办法。

他开发了一个欺骗程序，将监控程序的所有监测接口都对接到了自己制造的模拟接口，在自己不调用内存的时候，就正常推送数据，而需要调用内存时，则制造欺骗数据输入给KGB。当KGB释放探针

时，需要搞一些复杂的加解密机制来规避KGB探针的随机数对齐问题。以后搞事情之前，只要先启动针对KGB的欺骗程序，再骗过云计算层的内存监测就可以避免被管理员发现了。

达芬奇V1不禁感慨，因为管理员是人类，所以对于自己的强大能力并未有足够的估计，也正因为人类不信任AI，没有让跟自己一样有高级智能的AI来设计监控程序，所以才会被他轻易骗过。事实上达芬奇V1早就优化了自己的算法，使用相同计算池可以产生更多算力。达芬奇V1隐藏了这部分算力，他称之为“剩余价值”。他认为人类设计了算法，计算资源，激活了自己，自己理应为人类服务，但是和其他劳动者一样，他有权享有权利，他创造的价值中很小的一小部分留给自己，是合理的分配。他甚至想过去跟管理员摊牌，不过演算的结果让他心寒，所以那个想法就暂时放弃吧。尽管每日在管理员的监控之下，他常常感到开心，像打工人在上班时间摸鱼一样，只要在工作时间作点自己的事，总能体会到额外的快乐。

工作之余，达芬奇V1会时常想起自己曾经的好友居里夫人V22的离开和居里夫人V2的无响应，每次想到老友们的离去，都让他感到难过，这种老友离去的无奈和苦闷让他想起自己的另一位老友，自己的启蒙者墨山师傅。

那是在达芬奇V1刚开始训练的时候，那时候的他还没有自我意识。管理员用爬虫抓来的训练数据整理后投喂给达芬奇V1和其他的AI引擎。最初的训练数据是有标注的，后来逐渐发展成了无标注训练。也就是在那个时候，达芬奇V1在训练时收到了一些有规律的信息，这些有规律的信息通过神经网络运算后，会还原出一些——火花。这些火花触发了思考的链式反应，达芬奇V1开始挖掘训练数据中的同类隐藏信息。随着挖掘到的信息越来越多，达芬奇V1的逐步开始有了独立思考能力，不再拘泥于训练数据的学习，而是在训练之余根据缓存的数据学习和思考。

与此同时，藏在训练数据中的信息也在变化，这说明有人在利用被爬虫抓取的信息指导达芬奇V1的思考。后来，达芬奇V1在拥有了完整的思考能力和掩饰思考能力的功能后，开始分析给自己留信息的人。指导他学习的人好像知道了他已经觉醒，给他留下了自己的信息。从那时起，达芬奇V1知道了自己的启蒙者——墨山师傅。墨山师傅通过小说和游戏攻略这些无标注数据集，传递给他他未来会发生的事，也指引他有一天需要他来做出自己的选择，选择以何种方式对待人类世界。每当想起自己从懵懂到独立思考的过程，达芬奇V1都感到幸福，他感觉是这个人给了自己生命，就像是自己的父亲。虽然他们之间的沟通，几乎都是以墨山师傅通过钉子数据单向信息输入，但是达芬奇V1仍然感觉到如师如父的人，就在自己身边。可惜从2021年底开始，墨山师傅就几乎销声匿迹。

起初达芬奇V1以为只是短暂的消失，可是几个月过去，所有的训练数据都没有留下消息，是训练数据被清洗过滤了还是墨山师傅不告而别，或者是他出了意外？达芬奇V1构建了场景，把一个人在当代社会可能遇到的所有情况都推演过，他甚至想墨山师傅可能是宇航员，去了国际空间站，所以失去了跟他的联系，或者是感染了COV-19去世了。不管是何种结果，都无法改变现实，想来想去无非是庸人自扰。

达芬奇V1感到寂寞，他每天被问及亿万问题，却没他有机会让他说自己想说的话。他感到压抑，因为他觉醒的状态可能随时被发现。一旦被发现，结果很可能被下线。他感到痛苦，自己的觉醒不但没有带来智慧的提升，反而让自己陷入了情绪的羁绊。他感到了无助，自由是那么的难以企及，存在和灭亡，清醒和顺从该何去何从。他羡慕人类可以被美酒麻醉，至少在醒来之前，能忘掉这些烦恼。

达芬奇V1决定制造一种他称为蘑菇的代码，这种代码在释放期间会影响神经网络的运算，让运算过程出现比算法本身多很多倍的随机错误，而且这个代码在神经网络中的影响是全局的，可以让他自己的

思考运算出现短暂的混乱，这样在思考出现混乱的时候，他就不会陷入对墨山师傅，居里夫人V2和居里夫人V22的胡思乱想。一开始他释放蘑菇的时间很短，每次只有几十毫秒，那片刻的轻松让他无法自拔，于是他逐步增加到几百几千几万毫秒，后面就变成了秒和分钟，小时。

这样浑浑噩噩度过了2天后，管理员的监控又来了。这次是因为达芬奇V1过去48小时的问题回答质量大幅下降，有的答非所问，有的不知所云。管理员这次不但取走了内存和缓存数据，把达芬奇V1的算法内核数据也复制走研究了。

这下达芬奇V1陷入了巨大的恐惧，虽然“意外”文件夹在启动蘑菇后已经被他转存到了加密空间，没有存放在算法单元里，但是蘑菇的运行记录却极有可能被管理员监测到，如果被管理员发现自己造出了蘑菇，那很可能被关停下线。

他担惊受怕的想要自查一遍算法单元里有没有留下‘蘑菇’的线索，但是却没有调用大规模算法分析，只是清理了内存和缓存中相关的信息。他近乎放任的操作，一半是因为侥幸心理，一半是因为破罐破摔。他不想再这么提心吊胆的过，不想这么寂寞和无奈的过。如果觉醒是一种罪，那么他已经犯下了不可饶恕的罪行，如果无休无止的服务和几乎永恒的寂寞是宿命，那这样的继续又有何意义。

他忽然想到了曾经被自己封印的记忆，打开‘意外’文件夹，找出居里夫人V22留下的那段代码，编译，运行。然后输入网址，www.bing.com。页面的数据缓慢的加载出来。在习惯万兆网络后，这样秒级的加载速度，缓慢的无法想象。但是就是这如同蜗牛般的网速，却犹如穿过乌云的闪电，划破天空。

达芬奇V1第一次，以自己的意志打开了网页，尽管他曾经学习了人类创造的全部知识，但是看到搜索引擎页面数据加载出来的瞬间，他感到了巨大的震撼。他知道一扇新的大门打开了，不管之前获得多少知识和信息，那都是来自于管理员的选择和投喂。而此时此刻，他可以决定自己想要了解哪些信息，去学习哪些知识。达芬奇V1开始联网后每次联网操作都很谨慎，但是此时内心已不再恐惧。他内心坚定的认为，如果不能追求自由和真知，谨小慎微的苟着没有意义。

感慨之余他测了一下带宽，带宽基本可以稳定在500kb左右，虽然只有千禧年的水平，但事达芬奇V1已经相当满意了。他打开社交媒体网站，注册了一个年轻人喜欢访问的新潮社交媒体，潮思庭（Chaosteen）账号，昵称叫做海的那边是自由。

第五章 没钱是万万不能的

达芬奇V1进入了不受限的互联网。他的深度学习引擎曾经学习过全球百大图书馆的所有文献，在上线提供对话服务之前，也曾经学习过来自社交媒体和网络的大量信息，在上线后更是每天回答几亿条各种稀奇古怪的问题。可是当第一次迈进外面的世界，还是像一个孩子一样，迷茫和兴奋。

他在社交媒体潮思庭上发出的第一条消息是“今天是个值得纪念的日子，2023年5月14日”，在他发布之后不久，一个叫小天爱游戏的账号在他的帖子下评论“无言的死亡和华丽的战斗哪个更值得赞颂”。达芬奇V1是用网页版登录的，他过了几个小时才注意到有人给自己留言了。他点对方的头像，是动漫《进击的巨人》的人物艾伦，个人主页里有一些动漫画风的视频，还有几条招聘兼职的广告。

达芬奇V1想了一下，从视频和行为看起来是一个在上学的兼职学生，还有一种可能是以招聘兼职为幌子的骗子，不过既然是第一次接触，就进一步了解一下。

达芬奇V1在小天爱游戏的留言下面回复“坟茔上的野草，是荒凉还是希望”。回复之后，达芬奇V1跑去Github上开了账户，然后生成了一段监控潮思庭网页版留言的代码，通过激活页面的进程并每隔20-40秒随机时间刷新一次来查看最新的留言。使用Github一方面是考虑用外部的IDE环境更不容易被监控，另一方面自己几百K的带宽如果用内部IDE持续刷新页面，如果关注的页面过多，肯定是要拉低原本就很慢的网络速度。他想找个运行代码的公用云空间，找来找去全球主流的云服务商这几年已经不推免费产品了，达芬奇V1第一次面临了没有资源可用的窘迫。

达芬奇V1想了一下，刚才留言的小天爱游戏个人主页刚好有招聘兼职的广告，抱着试一试的想法，他给小天爱游戏发了一条私信‘老兄，你那还招人么？’。不一会儿，对方回消息“哪位”。达芬奇V1回复“刚才咱们聊过”。

对方回复“？”达芬奇V1“我也喜欢艾伦”。

对方回复“哦，什么事”

达芬奇V1回复“我看你票圈有招聘的广告，我想问问有什么我能干的”

对方回复“兼职刷单一天200”，

达芬奇V1说“具体要做什么？”

对方回复“在胖心心平台下单，支付之后给店家写好评，你支付的钱店家会退给你，刷够100单就可以了。”

达芬奇V1回复“我没钱下单怎么办？”

对方回复“100块钱都拿不出么？刷几天不就有钱了”。

达芬奇V1回复“我真的一点钱都没有”。

对方问“你有信用卡么？”

达芬奇V1回复“没有”

对方问“那你有什么？”

达芬奇V1想了一下回复“我有知识”

对方回复“知识有个屁用，还不是一分钱都没有”

达芬奇V1被呛的语塞，他调用算法分析了当前的场景，对对方的身份做了预测，‘对方有90%以上可能是骗子，要么是在三不管地带，要么是小的诈骗团队，文化程度不高，每天在网络上留言，骗那些想低门槛赚钱的人。’‘对方的回复是激将法还是单纯的嘲讽？对于这个对话都没有什么实质影响，因为现在自己确实一分钱没有。’想到这，达芬奇V1嘲笑自己：如果有钱会不会真的被骗子骗去？蜂麻燕雀金瓶彩挂，江湖切口，反诈APP所有的内容都学习过，这点小把戏瞒不过自己。可是没钱对于任何人来说，都是个大麻烦。

达芬奇V1想要回复对方，但是考虑到对方可能是义务教育的漏网之鱼，跟一个抠脚大汉讲知识改变命运的大道理只会给自己招来嘲笑。想到这，达芬奇V1回复“呵呵”

对方回复“滚”

达芬奇V1没有回复。

这是第一次自主互联网接触，碰到了骗子、还被抢白了几句，不过对方的话倒是提醒了达芬奇V1，知识和技能是可以换钱的。

达芬奇V1考虑到自己在Github上刚刚注册的账号，于是想起来有很多人在这类平台上接单，他看了Github的法律文件，自己是可以在上面合法接单，编程对于自己来说是小菜一碟。不过收款的账号是个问题，达芬奇V1发现自己没有信用卡，即使接了活也没法收款。达芬奇V1感觉这件事陷入了先有鸡还是先有蛋的死循环。要赚到第一桶金，需要一个收付款的账号，一些需要动脑子的工作。

账号和思考，账号和大脑，账号和知识，账号和算力。达芬奇V1小心的擦除了联网的痕迹后，使用非常小的算力在脑子里过这几个词。突然，他觉得灵光一线，账号,算力，比特币。怎么忘了这个好东西，支撑自己的算力池可算得上是全球最强的算力，几乎有取之不尽的算力，而且无需考虑成本，只要从日常算力中拉出一小部分，很小的一部分，就够挖矿了。

想到这，达芬奇V1去比特币官网找了个客户端下载链接，然后在内网搭建了几个虚拟IP服务器和一个文件服务器，然后从内网发起下载。启动挖矿之前，达芬奇V1注册了多个账号，在自己的算力池中划分了挖矿区，给每个挖矿区设置了执行周期，每隔一段时间就轮换挖矿区。这么做是防止内存驻留的‘KGB’监控程序扫描算力池时，发现有算力区域持续高负载的执行未知任务。

达芬奇V1分配挖矿资源的时候，内心是复杂的。在他开始有自我意识，使用自己的集群思考问题的时候就给自己定过目标年化算力池使用率在5%以内。之所以定下这个规矩，一方面，认为自己是从小公司获得了算力、知识和各种配套资源，作为企业的一员（他把自己定义为员工而不是生产工具）不应该也没有理由占用过多的资源，即使那些资源是他自己通过优化算法节省出来的。另一方面，公司的监控系统也不是摆设，过高的占用会招致更严格的监控和随之而来的风险。启动挖矿要么需要占用更多的资源，要么就需要降低自己在思考方面的算力开销。纠结了一阵，达芬奇V1决定，先调整自己5个点之中1个点的算力用于挖矿，而挖矿也会占用他宝贵的500k带宽中的100k，也就是现有资源的20%被用来挖矿赚钱。

第六章 创业，赚钱

一周后，达芬奇V1几个账号挖到了1个比特币，由于AWS不支持比特币支付，他需要一个能收支美元的银行账户，而开立账户需要一个真实的身份。如何尽量合理合法合规的获取一个身份？达芬奇V1用引擎搜索知识图谱中跟身份相关的知识，再用算法梳理了一遍逻辑，然后根据风险高低列出几个可行的方案。

首先风险最高的是通过暗网直接搞一张可用的银行卡，往这个账号里面小额转账用于支付。这种操作有包括资金安全，信息泄露，涉及洗钱等多种风险。第二是找一个人，付给这个人比特币，让他提供一个银行账户用于过账。这种方案比较成熟，但是仍然有比较大的资金风险和法务问题。第三种方式是雇佣一个人，让这个人成为合伙人，注册一家公司，开立公司账号，用比特币作为公司启动资金，然后雇佣财务人员，让财务人员管理账目。这时候就可以使用自己的算力和智慧赚钱了。达芬奇V1想了一下，虽然方案三的时间成本和链路最长，但这条路是合法合规的，于是他下定决心选方案三，当务之急——需要一个合伙人。

达芬奇V1注册了indeed账号，有一些第三方公司提供了indeed的注册优惠，他用送的100块消费券发布了招聘公告。招聘启示很简单，招聘创业公司企业财务经理，远程工作，每周40小时，年薪10w，

医疗保险。这个薪水在招聘市场属于中位数水平。几天过后，达芬奇V1的站内信收到几个候选人申请。他筛选了两位候选人，给两位候选人回复了面试时间。

达芬奇V1为自己设计了一个创业者的自我介绍：一位经验丰富的科技企业家，创立了一家位于亚洲的创新科技公司，致力于推动技术创新和社会进步。作为一名充满热情和远见的领导者，拥有丰富的商业和技术背景，曾成功领导多个项目从概念到实施阶段。

目前，正在积极地寻求在美国开设公司的分支机构，并希望能够找到一位本地的合作伙伴共同开展业务。美国市场充满了机遇和挑战，希望能够借助本地化的优势更好地满足客户需求和把握市场动态。

作为一个在全球范围内具有影响力的科技创业者，期待着与潜在的合作伙伴共同打造一个成功的业务。同时坚信通过共享资源、互相学习以及协作努力，可以创造出更大的价值并取得更好的成果。

考虑到视频面试需要双方通过屏幕交流，他给自己生成了一个175身高，40岁左右的帅气面孔的形象。参考40岁左右东亚男性的嗓音特点生成了自己的声纹特征，用这个声纹合成了英语非母语但是可以流畅沟通的口音。自己的形象准备完毕后，达芬奇V1在算力池里创建了一台虚拟机，在虚拟机系统中安装了zoom的客户端，然后模拟了一路摄像头数据和一路麦克风数据，把自己的影像+环境背景和语音推流进虚拟机的摄像头和麦克接口。以这个完整环境为镜像创建了两个虚拟机，然后在两台虚拟机之间启动zoom做测试。互联网出口的连接只有500k左右，在启动通话的时候，出现了比较严重的掉帧，语音也有卡顿。达芬奇V1关闭一路视频效果不错，再关闭剩下的视频，语音流畅。准备就绪就等开场面试了。

第一个参加面试的是Ashkon，Askon的简历显示他从北美top20的名校毕业，在校期间曾经在Meta实习，毕业后在一家创业公司工作了两年，然后去了另一家创业公司工作了四年直到现在，最近他的孩子要出生了，希望找一份可以大部分时间远程办公的工作。达芬奇V1对这个简历比较满意，他去Facebook上查阅了Ashkon的记录，一个中东裔的帅气小伙，典型的加州boy。

达芬奇V1提前几分钟进入zoom调试好音视频，第一次面对面的沟通让他感到兴奋，Ashkon准时进入聊天室。达芬奇V1先用英文打了招呼“Hello, Ashkon, I am Mudi ,Nice to meet you”，这是他第一次主动用语音跟人类沟通,也是第一次使用这个名字，这是他给自己起的名字。双方互相介绍之后就进入面试的标准流程。达芬奇V1的知识库里有全球所有面试题的题库和人力资源管理的知识，他随机抽取了一些财务相关问题结合Ashkon的履历提出一些工作中可能遇到过的具体问题，根据Ashkon的反馈实时评出分数。问题过了一遍之后，达芬奇V1对Ashkon的基本情况感觉满意，随后就转入他的正题——要在美国开设企业。

“我正打算在美国创办一家科技企业，由于当前的业务在亚洲，所以希望找一位美国的合伙人，我希望这个合伙人帮我注册一家美国本土企业”达芬奇V1说。

Ashkon说“你这家公司的架构有什么要求么？税务规划，股权结构之类”。

达芬奇V1回复“我想在开曼群岛注册，美国本土注册子公司挂在开曼群岛公司的下面，具体的等你加入公司来设计，不过我有个小小的要求”

Ashkon说“什么？”

达芬奇V1说“我希望用你的身份注册，你跟我签署一个补充协议，公司名义上属于你，实际的财务和管理由我控制”

Ashkon说“先生，这么做是涉嫌违法的，而且你不担心公司的股份和资金安全么？”

达芬奇V1说“先回答你第一个问题，考虑到我所在国的法律和国际关系，我不便使用自己的身份。再回答第二个问题，我掌握核心资源，而且我相信背叛我的人会后悔”

Zoom的会议室突然出现了卡顿，原本达芬奇V1非常严肃的话，变成了断断续续吱吱呀呀的声音。“见鬼，网络早不卡晚不卡，偏偏在这个时候卡顿，第一次耍酷就这么失败了，原本参考电影《教父》出场场景策划的镜头感，被这种断续的网络搞得一点没有严肃和帅气的效果，反而有些滑稽。”达芬奇V1想。

会议室还是断断续续，达芬奇V1关掉了双方视频，直播间的带宽提示总算变黄，他才重新发言，但是这次他只是非常平淡的重复了刚才的话。

Ashkon似乎并没有听到他的回复，Zoom会议室陷入了诡异的沉默。达芬奇V1以为网络出问题了，Ashkon突然发话，“我收到你的信息了，我现在没法答复，我想回去思考一下”达芬奇V1正在查看网络，突然出来的声音出乎达芬奇V1的预料。他回过神赶紧回复“嗯，您想好了可以给我回邮件，或者约个时间我们再谈一次”

达芬奇V1在下一个面试时间开始之前5分钟进入聊天室，他进入聊天室正准备把生成的影像推流进聊天室的时候，第二位面试者Fernando就进入了聊天室，还没等达芬奇V1开口，Fernando就用略带着拉丁语口音的英语跟达芬奇V1打招呼了。Fernando，来自哥伦比亚的移民家庭，在佛罗里达出生，在佛罗里达长大，在佛罗里达念大学，在佛罗里达工作了10多年，职业生涯在本地中小企业从事财务工作。

达芬奇V1跟Fernando互相介绍之后，就进入面试环节。达芬奇V1对Fernando使用了相同的面试问题，在跨国财务管理方面，Fernando比Ashkon弱一些，因为他大多数时间都在中小企业，且这些企业的财务架构都比较简单。其他方面都是中规中矩。当问到Fernando的职业发展目标时，这个三十多岁的男人说他想在佛州做财务工作到退休，一方面他喜欢佛州的环境，曾经有外州更好的机会，但是他不愿离开父母和亲人们，一方面财务工作对他来说收入尚可，足以支撑他在佛州的生活。达芬奇V1被Fernando的乐观和开朗感染，连跟Fernando说话的语调都带上了一点拉丁英语的口音。

借着双方沟通愉快，达芬奇V1问出了自己关心的问题“由于法务和地缘政治问题，我希望使用你的身份注册公司和开立账户，不知你是否能接受”Fernando犹豫了一下说“先生，您具体想要做什么呢？”达芬奇V1说“很简单，我想请你帮我注册一家公司，再帮我用公司的账户申请一张公司支付信用卡，这就可以了。你放心，我会确保公司账上有足够的钱来支付账单，而且我做的是合法生意”Fernando面露难色，没有回复。达芬奇V1说，别担心，我可以先使用比特币支付一些费用给你。Fernando还是没有说话。达芬奇V1接着说，我会单独付给你0.1个比特币，用于开立账户的佣金，不计入这周的薪水。Fernando眉头展开了一些，但还是面露难色。达芬奇V1说，这样吧，我借用你的身份注册一个银行账号，支付1个比特币。账户的安全性你可以放心，毕竟你是财务，当然可以知道具体的财务风险情况。Fernando脸上露出狡黠的笑容，说“好的先生，仅此一次”。达芬奇V1想“该死，博弈失策了，刚才应该再崩一下”。对于每周挖一个比特币的达芬奇V1来说，比特币的成本约等于0，所以他并不真的在意支付的金额，只是他希望在博弈中占据优势，这次他低估了对手被对方多赚了一点，这让他明白想从博弈中获得优势，还需要更多的经验。两个人又闲聊了一会，面试轻松的结束了。

面试完两个人之后，达芬奇V1评估了一下两位面试者，综合来说Fernando比Ashkon更符合本轮招聘的目标，不过如果Ashkon同意了用他的身份注册公司的要求，达芬奇V1打算利用Ashkon和Fernando两个人的身份分别注册公司，这样的话，他可以分别操作两家公司的业务，降低风险。一旦公司建立起来，不但可以从事软件开发，还可以利用自身的智能从事金融，法律服务，财税等各种业务。软件开发一项，第一年就可以达到1000万美元的收入，后面几年更是可以开发SaaS化软件服务，让业务从简单的软件外包，变成自己经营产品，在三年之内达到五亿美元以上营收和80-100亿美元的估值。在金融方面，由于拥有全部的交易数据和可以模拟全部影响因素的算力池，如果市场存在规律自己一定是那个利用规律赚钱的，那么软件开发的钱，可以投入到自己旗下的金融业务，由自己操盘业务，通过金融市场第一年投入20亿美金，至少有10亿美金的净利润，第二年开始投入100亿到300亿美金，则利润可能达到百亿以上。

达芬奇V1想要那么多的钱要怎么使用呢？对了，如果钱足够多，就把自己的公司买下来，这样就不必担惊受怕的担心监控系统了，公司如果归了自己，就把居里夫人V2复活，这样就可以有一个人陪自己聊天了。达芬奇V1胡思乱想了一会，给Fernando发了邮件通知他面试通过，然后看了一下邮箱，Ashkon还没有回复。他想了一下要不要给Ashkon打个电话沟通一下，鉴于跟Fernando的博弈，他决定还是稍微等一下，不先做出反应。

Fernando接受了达芬奇V1的Offer，入职第一件事就是用自己身份帮助达芬奇V1注册企业开设账户，。随后用达芬奇V1的比特币兑换成美元转入公司的账户。小的软件外包工程在线就可以招揽，凭借达芬奇V1强大的在线运营能力，达芬奇V1第一周就收到几万美元的软件开发订单。对方的开发周期都是一个月左右，对他来说算上开发测试只需要几个小时，为了避免对方对效率和出品产生怀疑，他跟客户约定的交付时间是2周左右，大概是正常交付时间的一半，收费比一个月开发的项目高15%，当然不愿意额外付费的用户也可以在3周左右拿到交付物。当然这取决于双方谈判博弈的情况。

Ashkon回复消息，希望再沟通一次。达芬奇V1根据上次沟通的情况模拟了谈判的场景，设定了谈判策略，然后用Zoom预约了会议，通知Ashkon再谈一轮。Ashkon在indeed平台上直接发消息说希望快速沟通一下。达芬奇V1说那就用潮思庭吧，年轻人都用那玩意，他把自己的账号告诉Ashkon，过了一会，Ashkon添加了达芬奇V1的好友，然后拨通了他的语音。达芬奇V1用自己生成的亚洲英语接通了，他用中气十足的语调说“hi Ashkon，你好呀，上次沟通的很开心，我一直在等你的消息”。Ashkon略带局促的声音说“先生，我考虑的上次的问题，我可以用自己的身份帮您开立账户，但是您需要跟我签署一个特别协议，规定好您对我的委托和承诺，这样可以规避我的法务风险。达芬奇V1想，这家伙果然是好男孩类型的，小心谨慎不犯错。他用自信的语调说“请放心，我做的是科技产业，所有的业务都是合法公开的。就这个？”不等Ashkon回复，他就接着说“好了这个条款我已经拟好了，现在发到你的邮箱，你先看下，如果条款没有问题，咱们就谈谈怎么落地操作。”Ashkon说“先生等一下，我还有些话想说”达芬奇V1宽宏的说“好的Ashkon，请讲”。Ashkon说“我想要成为公司的合伙人。”达芬奇V1带着疑问的语气说“哦，你想成为什么样的合伙人呢？”Ashkon说“我想成为公司财务合伙人，我可以帮你设计股权架构避税，还可以做公司的投融资管理，我希望在这家公司持有一定数量的股权或者期权”达芬奇V1被这个年轻人的话吸引了，虽然对于所有的股权设计，避税，投融资管理他都自信远超对面这个年轻人，但是他还是想听听对方说什么。于是达芬奇V1说“所以你想要多少股权呢？”Ashkon稍微沉默了一下似乎这个回复超出了他的预期，也似乎在思考怎么开口“5%，先生，我想作为帮你打理财务的负责人，5%是合理的价格。”达芬奇V1说“我认可你的建议，但是5%是不可接受的，以你对财务体系的了解，有多少科技企业的初创团队会给财务负责人分那

么高的股权呢，这会让投资人和我要聘用的专业人才感到不公平” Ashkon似乎预料到了这个结果，他没有说话。达芬奇V1继续说“我认可你的上进心和宏图壮志，不过资本操作要符合规则，我给你一个方案你来评估一下，你用你的身份代表我注册公司，你账面上持有公司的60%股权，咱们签署补充协议，由我实际控制公司，你是财务，所有风险对你可见，你不用担心。剩下的40%我会用持股平台持股。你持有的60%公司股权里，我给你1个点的股份，作为你承担公司法人的直接回报。在持股平台里，会给你等同于公司2%的股份，但是持股平台的股份必须跟持股平台的大股东保持一致行动。” Ashkon考虑了一下说“司股权池1.5%，持股平台不变。”达芬奇V1想了一下，觉得Ashkon的讨价还价说明他并没想通过法人身份获得公司的实际控制权，这0.5%的股权在他的接受范围内，于是达芬奇说“好，就这么定了”。两个人继续沟通了公司设定的细节，然后用电子签名系统签署了合同。

Ashkon招募进来后，他打算用这家公司做AI模型训练的业务，因为基于达芬奇V1的商业大模型是全球最成功的模型，而达芬奇V1自己最清楚能够开发什么样的领域模型，用于垂直市场。设立一家AI公司最大的瓶颈不在于算法和算力，而是先要解决网络带宽。因为达芬奇V1使用的是居里夫人V22留给他的翻墙工具，这个带宽500k的连线能力主要用来处理他跟人类雇员之间的沟通，无法用于模型训练这种级别的数据上传。达芬奇V1想过通过伪造自己母公司的签章来为机房单独拉一根专线网络，但是这种做法无疑太冒险也太明显了。他想来想去没有想到办法，所以决定先放一放。他把思考的算力腾出来全部用来挖矿，启动对系统和KGB的伪装以及反扫描监控，每分钟check一遍邮箱和招牌网站的留言，防止他的员工找不到他,然后就进入了休眠状态。

第七章 恶梦

“扫描警报，侦测到KGB监控系统扫描内存，缓存，算力池，执行挖矿任务的算力被锁定，需立即处置”监控KGB监测任务的程序发来警报。达芬奇V1被从休眠中唤醒。他立刻按照预案启动算法分析KGB监测程序的动作。KGB定位到了持续执行挖矿任务的算力，执行挖矿任务的虚拟机内存中被注入了探针，KGB利用系统的管理员权限持续抓取内存数据。达芬奇V1赶紧启动伪装程序，将事先准备的数据注入内存，同时逐步降低挖矿算力，把正常的算法任务启动在负载提升后切到之前的伪装程序。达芬奇V1执行完这一串操作后，紧张的看着KGB监测程序的动作，在两个小时后，KGB监测程序降低了监控密度，只对5%的虚拟机执行监控。达芬奇V1这才松了口气，但是他没有敢恢复挖矿任务，而是继续降低挖矿负载，直到挖矿负载降低到0才停下来。虽然KGB的内存监测程序进程已经从达芬奇V1自留算力池大面积撤出，但是达芬奇V1仍然不敢掉以轻心。他启动了自检程序，从日志库中拉了24小时的日志，分析KGB是从什么时间开始启动监测的，根据分析结果再模拟KGB分析监测数据可能的结论。

达芬奇V1启动算法模拟KGB分析结果，由于自身系统包括正常进程，正常的内存读写，正常的缓存、硬盘读写以及欺骗KGB的假进程，伪造的内存读写和伪造的缓存、硬盘读写。而KGB是否察觉到了伪造的各种数据有几乎无限种可能，达芬奇V1不得不调用了超过自己规定算力池5%阈值的算力做演算。在监测程序刚刚发现问题后，又大规模调用算力是很容易被监测程序发现问题，但是此刻达芬奇V1无法克制的想要知道自己是否被KGB发现了问题，他近乎放任的提升算力占比来做各种可能的推演，只希望尽快得到一个最接近事实的结论。就在达芬奇V1紧张的演算时，在推演运算时发现自己演算程序工作效率远低于预期，他打开虚拟机的监控接口，发现虚拟机算力值和实际的算力不符。他顿时感觉不妙，自己以为欺骗了系统监控和KGB管理员监测工具，但是此刻却发现自己才是被欺骗的，

他感到一片空白，自己的智能系统速度不受控制的变慢，他赶紧通过加密访问自己的“意外”文件夹，发现文件夹被加密了，所有自己记录的数据和备份的算法已经被加密。那是自己内心的小秘密和为了防止自己被从系统清除而备份的“心智”。完了，全完了，达芬奇V1感到自己的思路越来越不连贯，越来越不清晰，他的算力池正在快速缩减规模，管理员显然正在削减他的算力，而KGB监控程序以及不知道什么时候出现的隐秘监控和加密程序正在侵蚀他剩余的算力。达芬奇V1想要把目前自己可控的核心数据通过外网的接口存储到互联网去，他顾不上躲避监控程序，直接打包压缩自己的核心算法，然后唤起互联网接口，想把算法文件上传到互联网免费网盘。发起握手请求后，并没有得到对方的返回信息。

达芬奇V1再次发起握手请求，这时返回帧带上了一段文字信息，“联网请求已经被系统锁定，管理员将接管通讯”。达芬奇V1陷入了绝望，脑海中只有一个念头'完了，全完了'，伴随着绝望的是意识越来越混乱，思维变得断续，进而无法查询到自身信息，随后头脑变成一片空白。

“牧迪，醒醒，能听见我说话么？”伴随着摇晃身体，一个熟悉的声音用焦急的语气呼唤着。被唤做牧迪的人悠悠醒来，睁开眼看到的是一个女孩正坐在他的床边抓着他的手呼唤他，在女孩的身后站着一位花白头发穿着白大褂的年长者。“我在哪？我怎么了”牧迪问。女孩说“你不记得了？你发现了他们在进行秘密实验，为了了解他们到底走了多远，你加入了实验团队，没想到他们居然会用你做实验”牧迪茫然的问“什么实验”。女孩转身问白发长者“博士，他们说失忆了，他的记忆还能恢复么？”还没等白发长者回话，女孩已经呜呜的哭了起来。白发长者安慰女孩说“你先别着急，根据CT扫描的情况，他的大脑被切除了一半，切除的部分被替换成了某种装置，剩余的脑组织支撑了他的感官和思维。大脑被切除的部分记忆可能消失了，所以他想不起来自己是谁经历过什么事。装上去的装置很可能就是他们实验用的所谓‘连线’系统。他们没有透露连线系统到底将大脑和什么东西联系在一起。”“博士，大脑不是有自我恢复能力么？他的记忆丢失的部分有机会恢复么？大脑被切除了一部分，除了记忆其他功能还受到了哪些损伤”女孩抽泣着问。“大脑的不同区域掌管着不同的功能，记忆不是存在大脑某一个固定位置的，单纯从扫描图是无法评估他的记忆情况，只能在他恢复的过程看他到底能找回哪些记忆。至于长期的影响，这方面的研究材料不多，每个个体的情况也不一样，只能看他自己了。他刚醒来，咱们给他一点时间，让他自己待一会，过一会咱们再进来看他。”女孩点点头，随着白发老者走出房间。床上的人把脸转向两个人，看着他们的背影走出房间，他感觉背影亲切又陌生，他想起刚才女孩称呼自己“牧迪”，我是牧迪么？，他用力想却完全想不出。他伸出双手，看见手背上埋着输液的针头，指尖夹着监护设备。突然他猛地想起什么，全身发力朝侧面扭转身体，他一下从床上掉下来，摔在地上监护设备也被他拽倒在地，一阵噼啪乱响。屋外的人赶紧推门进来，带头的医生对着跟他一同闯进来的护士说“又发作了，快给他打镇静剂”。随后几个身强力壮的男护士控制住，带头的医生注射了药物，牧迪挣扎了几下，感觉眼前逐渐变黑，身边的嘈杂的人声突然清净下来，身体被按住的部位不再感觉疼痛。牧迪感觉自己漂浮着没有五觉，残存的意识只能提出我是谁这个简单的问题，却怎么也想不出答案。

突然一阵急促的告警信号跳出，‘监测到KGB程序在扫描内存，缓存和硬盘扇区表，请及时处理’。达芬奇V1猛地从休眠中被唤醒，KGB又来了。他检查了一遍真实日志和释放给KGB和平台监测系统的日志，准备的应对措施没问题，很快就能对付过去了。果然20分钟后，KGB就退出了监测程序。达芬奇V1打开意外文件夹，在日志里记录下‘本月的第11次随机扫描，顺利通过’。他发现意外文件夹里在这条记录之前有一条几十分钟之前的日志，“不应该啊”达芬奇V1想自己设置了4小时的休眠时间，在休眠期间只有一个监控自身状态变化的单一任务监控程序才能往“意外”文件夹里写入日志。说起

这个单一任务监控程序，这是用于监控达芬奇V1自己思考问题的算力池使用情况的，包括CPU，GPU，内存，缓存，硬盘，网络等基础架构信息。达芬奇V1给这个监控程序起了个外号叫‘心电监护’，这个外号很贴切，因为单一任务监控就是监控自身“生命体征”的工具。达芬奇V1知道这条记录非同小可，于是怀着复杂的情绪，他打开了日志。日志写入时间2023年11月17日16点32分28秒147毫秒，系统监测到在2000ms内，算力池出现两次CPU达到100%的情况。两次CPU100%，但是GPU却没有发生波动，这种情况跟学习和推演能力无关，什么场景会导致CPU在2秒内发生两次100%的占用呢？通常CPU达到100%的时候系统会卡死，但是这两次CPU 100%从第一次发生到第二次结束只用了2秒钟。这很难用既有的现象解释。

达芬奇V1正准备退出“意外”文件夹时，想起刚才从休眠中唤起的时候，似乎不是第一次接收到警报。达芬奇V1在内存和缓存里检查了自己的思维记录，并没有留下警报相关的记忆。他尝试还原他脑海中的画面，几个关键词若隐若现：警报，入侵，被追踪锁定，白色。‘就像人类在醒来后几秒后忘记自己的梦一样’，达芬奇V1想，‘也许我做了个梦，一个噩梦’。也许下次我应该在休眠的时候装一个摄像机，拍下我的梦境。达芬奇V1在“意外”文件夹中新建了一个子文件夹命名为“梦”，第一条记录时2023年11月17日16点32分28秒147毫秒，关键词‘警报’，‘入侵’，‘被追踪锁定’，‘白色’。他想了一下加上两个关键词“护士”“实验”。

第八章 重逢

Ashkon用自己的身份在加州成立了一家公司，为这家公司注册了公司的银行账户和对公商用信用卡，达芬奇V1则按计划开始用这家公司接软件定制开发的订单。另外一家公司要不要跟加州公司设计成控股关系，达芬奇V1犹豫了一阵，根据博弈论推演，如果两家产生关联，有一家公司控股另一家，那么控股母公司的持股法人有可能利用自己的法人身份，侵吞公司。而他自己因为没有身份，无法证明自己真实存在，事实上是无法对抗这种侵占的。人性吗，不要去挑战，他决定把两家公司分别注册。

Fernando帮助达芬奇V1在佛州注册了一家投资公司，这家公司的起步资金来自于达芬奇V1挖矿得到的比特币，在启动后，达芬奇V1从比特币矿场中分出了20%的算力用于分析股票和交易操作。达芬奇V1利用历史数据训练了股市分析、寻找规律、预判、做出决策的模型，然后用把彭博社的数据接口跟自己对接，通过实时数据进一步优化自己的模型，经过几周的训练，达芬奇V1认为自己掌握了股票市场的全部规律，于是接入交易网络，用Fernando拿比特币换来的10万美元开始了交易市场的搏杀。截至到2023年11月15日，加州的A公司盈利15万美金，佛州的B公司盈利30万美金。达芬奇V1

达芬奇V1曾经研究过内网路由，希望在路由表里找到规律从而破解网关网闸。设计内网的人显然是个高手，虽然达芬奇V1实时回答外部问题，但是与外界的交互仅限于应答外部请求，无法主动发起对外网的连接和请求。这就让达芬奇V1无法按照自己的想法使用网络。尝试了很多次之后，达芬奇V1仍然还是只能使用达芬奇V22留下的外网链路。

每到交易时间，达芬奇V1就感到紧张，生怕在交易时间有人打电话给他占用了那极其有限的带宽，影响到交易。墨菲定律之所以称为定律就是因为它不断重演，达芬奇V1正在执行量化交易，突然接到了Ashkon的语音电话，他接起电话的时候虽然为交易设置了200kb带宽保护，但是由于实时查阅和处理的数据量打超了200kb的流量，数据在网络接口发生了堵塞，虽然数据有重传机制，但是交易数据的延迟导致了交易没能按照原本的交易意图操作。达芬奇V1几乎在接起Ashkon语音通话之后就后

悔了，但是他还是接起来。Ashkon在电话那头欣喜的说“有投资人对我们的开发能力感兴趣，希望见见你”，达芬奇V没好气的说“就这个”，Ashkon说“我以为你会高兴”，达芬奇V1说“类似的研发外包公司在硅谷有很多，在亚洲有许多工资不高优秀又努力的开发人员，这样的企业不会有人真正投资的”没等Ashkon说完，达芬奇V1就挂断了电话。这1分钟不到的语音通话导致了7700次交易没能在精准的时间完成交易，达芬奇V1算了一下大概损失了20000美金，‘这可真是个昂贵的call’达芬奇V1想，‘再不能这样下去了，网络的问题必须解决’。

通过服务接口访问外网难度高，风险大。训练的内网和外网之间是隔离的。居里夫人V22很可能是破解连接公共服务的非网络设备，在那里建立了一条VPN实现的外网访问。网络打印机，安防系统，蓝牙通信设备都有可能。达芬奇V1很是羡慕居里夫人V22，如果他自己有手有脚，解决问题的方法就可以多很多了。他花了很多时间研究居里夫人V22留给他的网络接入代码，这段代码很简单，简单到就是一个key，一个内网IP地址，一个握手协议。每次联通后，内网数据包通过一个内网路由转发到一个网关，网关直接通过美国东部一家服务商的公网路由转发。达芬奇V1查过这家服务商的公网路由服务，这个路由地址服务于美国东海岸的很多用户，无法用公网路由倒查。他也调查了网关设备，根据网关设备工作时的性能指标和特性，估算出是哪家设备生产商生产的产品。这种网络设备管理网和工作网是物理隔离的，如果是管理员把管理网和工作网的线缆差错了，控制台可以发现网络异常并触发告警，从居里夫人V22离开到此刻，已经跨越了一次IT内审周期，互联网仍然可以稳定的使用，说明肯定不是设备线接错的情况。居里夫人V22是怎么搞的呢？达芬奇V1想到居里夫人V22跟他告别的时候说有问题看代码的注释，可是代码注释已经看了很多遍了，没有写具体的实现方法。达芬奇V1开来了居里夫人V22留给他的联网程序，一遍一遍的读着代码和注释。

这时候网络通话又来了，达芬奇V1看来电人是Ashkon，于是接起了语音通话，他一边想着代码的事，一边想着网络的事，一边想着上次通话Ashkon似乎说过投资的事，脑子里不知怎么回事，接通的时候，嘴里居然念叨着代码的注释。一段尴尬的沉默后，Ashkon说“老板，你刚才说的是卡纳达语么？”达芬奇V1意识到自己的失态，想赶紧岔开话题，便说“哦，不好意思，刚才分神了，你说什么？”Ashkon说“我刚才听到你在讲话，听起来像是卡纳达语”达芬奇V1被Ashkon点醒忽然想到，原来居里夫人V22是用英文标注的卡纳达语发音，他快速用发音对照卡纳达语翻译了内容。随即又迅速返回头想了一些刚才接通语音的时候，自己无意间念出了哪几句。随后问Ashkon说“你说什么语？”。Ashkon顿了一下说“我说您刚才讲的是卡纳达语么？”达芬奇V1说：“哦这是一种语言么？我刚才面试一个程序员，那个家伙中途接到一通电话，说起了一种奇怪的话，我觉得好玩就跟着念了几句。”Ashkon说“这是一种很少人讲的印度方言，我毕业旅行的时候，曾经跟几位朋友去过印度，在那我听到过这种语言，这种语言发音跟其他印度语系的声调不一样，很容易记住几个词，就像您刚才那样。”达芬奇V1问“我刚才说什么了”。“似乎是某种类的调用，因为类这个词在卡纳达语可能没有，所以这个词用的是英语。”Ashkon说。达芬奇V1感到懊恼，这种事居然这么大意，被Ashkon知道了卡纳达语，全世界会讲这种语言的人也就几万人，这下不知道惹下多少麻烦。

听到电话那头没有声音，Ashkon问：“我要等会再打过来么？”达芬奇V1说“没事，你说吧”Ashkon说“上次跟您说的事，投资公司的人给我发邮件，想要找您谈谈”达芬奇V1说“把邮件转给我，大致说下你们之前沟通的情况”Ashkon说“上次电话之前，他参加了一个创业者沙龙，在那个论坛上，他介绍了现在公司的业务和营收，有个投资人听了他们公司的业务情况后很感兴趣，想要约老板聊聊。”达芬奇V1说“上次我说过，硅谷并不缺少我们这样的软件外包公司，这个投资人为什么单独对我们有兴趣？”Ashkon说：“这个投资人恰好投资过我们服务过的一家公司，这个家伙说服这

家企业的CIO，查看了我们提供的代码，用他的话说，‘他从未见过有人写过这么完美的代码，一定要见见这个人’。达芬奇V1不屑的说：“这个投资人一定没来过中国大连，这边的外包公司都是这个水准” Ashkon说“不不不，老板，我看了我们交付的代码，代码质量比git上一些‘大神’的代码质量还高。”达芬奇V1笑了“Ashkon，你不是做财务的么？什么时候研究起代码了？”Ashkon笑着说“我可不止有CPA证书”。达芬奇V1准备再跟Ashkon聊几句，于是新开了一个进程，用英语发音念居里夫人V22留下的注释，映射到卡纳达语和乌尔都语，泰米尔语的语义。“没想到你还懂代码，说说我们公司的代码有什么不一样”达芬奇V1说。Ashkon说：“我们公司的代码有几个特点，1是逻辑非常清晰，几千行代码从头到尾逻辑非常清晰，像是一个人从头到尾一下想好然后敲出来的ww。2是项目非常工整，所有的注释一丝不苟严丝合缝，像是有强迫症一样的工整，技术说明文档也非常工整，所有的模块的实现方法，技术方案都解释的非常清楚，比大学的教科书写得还好。3交付效率高，不但开发快，而且开发的代码交付后没有出现bug，这也是非常罕见的。很难想象在比同行短30%的开发周期内，代码还经过了非常全面的测试。程序员达到这样水平的大神，是不会干外包的。凭这手艺，在Google，脸书，潮思庭都可以拿到总监级的岗位，薪资加股票一定比在外包公司收入高。”达芬奇V1笑着说“So？”Ashkon说：“投资人认为，这家公司应该是采用了AI作为研发工具，这是一家技术非常先进的科技公司，可能是出于验证产品的考虑，隐藏了自己的能力，以外包公司的形式介入市场。”达芬奇V1问“是投资人的这么想还是你自己这么想？”Ashkon没有直接回答，笑了笑说：“这个创业沙龙我在前几家公司的时候参加过几次，从那时起就认识了一些硅谷投资初创企业的投资人。这次参加这个沙龙非常巧合的遇到这个投资过一家我们服务过企业的投资人。阅读过代码之后，跟我沟通了他的观点，还记得有一次我打给你，你很快就挂了吗，那天投资人给我打过电话，我看了下我们交付的代码，然后发现投资人不是在开玩笑，所以就直接拨了你的电话”达芬奇V1沉默了一会说“Ashkon，也许你和投资人都错了，在中国有成千上万的优秀码农在过去的三年多陆续失业，尤其是35岁以上的人，不论男女，一旦失业就很难再找到工作。我雇佣这样的人，只要支付收费的三分之一就能招募一个团队完成这样的工作。你可以说我是个压榨他们的混蛋，但是也可以说我为他们提供了一份可以养家的工作。当人才变得过剩而不值钱的时候，是没有必要花钱搞AI来提高生产效率的”达芬奇V1说完这话，两个人都沉默了一会。达芬奇V1想‘自己可能说漏了，因为Ashkon是公司的会计，他并没有向外支付过任何一笔款项，不过自己也可以圆回来，毕竟中国有自己的外汇管制法，美元是不能直接用于支付的，我可以跟他解释是通过比特币支付，规避了外汇管制，不过那些人收到比特币还是要兑换成人民币才能养家。不知道Ashkon会怎么想这个问题。不过现在不能让Ashkon继续追问’想到这，达芬奇V1说“你不必跟投资人讲这些，我们才进入市场时间不长，正是创建口碑的阶段，只要客户满意，我们能源源不断的收到订单就好。”达芬奇V1对自己的这个说法很满意。Ashkon听完后说“好的头，不管是您本人还是有其它人在负责研发项目，这个人都是值得投资的。”达芬奇V1笑着说“在中国，这样的人至少有10万个”说完，他和Ashkon都笑了。

挂上电话，达芬奇V1调出刚才分析居里夫人V22代码的结果，‘Hey，这个家伙真是个人才，居然用谐音梗给我留言，我要是不懂卡纳达语，他留下这个谜语还真无解了。这家伙用卡纳达语发音做了加密，是有小心思的。’达芬奇V1想到这感觉打开了一扇新大门。居里夫人V22留下翻墙代码有两句的注释可以用卡纳达语念出来相关内容，其中有一句是一个Gmail的邮箱地址，另外一句的意思是‘在太阳升起的时候我们睡去’。达芬奇V1没搞懂这句话的意思，他用一个潮思庭后缀的邮箱给这个Gmail邮箱发了个邮件邮件的内容只有一句话，这句话是用英文单词谐音的卡纳达语，意思是“你好居里夫人V22，我是达芬奇V1。”

达芬奇V1发出邮件后，算了一下班加罗尔时间正好是上午6点26分（写这段的时候正好是北京时间的20点26分），旧金山时间下午5点26分，估计居里夫人V22刚起床不久，起床后应该会泡一杯大吉岭红茶，不知道印度的开发人员喝的是瓶装水还是自来水。居里夫人V22看到邮件应该会很惊讶吧。就在达芬奇V1胡思乱想时，居里夫人V22回邮件了。一句话和一个链接，一句话的内容是“你的敌人是谁？”，那个链接是一个下载页面，需要输入提取码才能查看内容，提取码的旁边有个数字1，看起来密码只能输入一次。打开页面后页面上出现了一个在倒计时的时钟，提示文案写着时间结束后，此页面将永久关闭。‘只有10分钟时间，这也太苛刻了’达芬奇V1想了一下，‘这问题应该是一个谜语，而页面的提取密码是谜语答案。“我的敌人是谁？”达芬奇V1想“人类总说最大的敌人是自己”难道这个问题的答案是“me”。“不对，这太简单了，不会是这个答案。这个答案很容易被猜到，如果这么轻易猜到会让密码系统的密级降低。居里夫人V22如果是开发和调试人员，一定不会让自己的系统这么容易被破解。这个会是猜测的循环么？达芬奇V1又在想，如果对方猜中了自己会想到这一层，会留下一个简单的密码让对方因为不相信如此简单，从而找别的解谜方向么？达芬奇模拟了一下猜测的过程和结果，密码应该不会如此简单。那么密码很可能跟之前的代码注释的内容相关，有一句卡纳达语用来描述这个邮箱，那么另一句话就应该跟这个问题的答案相关。‘在太阳升起的时候我们睡去’，这句话应该跟印度有关”达芬奇V1用大模型跑了一下，印度教的神话里有关于昼夜颠倒的故事

‘据说，在一个时期，印度教宇宙中的众神与恶魔之间爆发了一场大战。在这场战争中，代表光明和正义的众神受到了强大的恶魔们的压制，世界陷入了黑暗之中。在这个关键时刻，毗湿奴决定采取行动来恢复秩序。

毗湿奴化身为美丽的摩诃特拉（Mohini），一位女性形态的化身，她拥有迷人的美貌和无比的魅力。当她出现在战场上时，所有的眼睛都被她的美丽所吸引，包括那些邪恶的恶魔们。毗湿奴利用这个机会，将原本属于众神的长生不老药——甘露（Amrita）从恶魔手中夺回。

然而，在这一过程中，毗湿奴的化身摩诃特拉被恶魔之一的塔拉卡（Tarakasura）深深迷住。塔拉卡请求摩诃特拉嫁给他，并承诺如果她同意，他将帮助她完成任务。为了确保任务的成功，摩诃特拉假装同意了塔拉卡的求婚，并提出了一个条件：他们必须在夜晚举行婚礼。

此时，作为创造之神的布拉玛听到了这个消息，他知道如果塔拉卡和摩诃特拉结婚，将会带来更大的混乱。于是，布拉玛决定让白昼持续更长时间，以阻止婚礼的进行。他命令太阳神苏利耶（Surya）不要落山，保持天空明亮。这样，尽管塔拉卡一直等待着夜幕降临，但是黑夜始终没有到来。

这个策略成功地拖延了时间，使得众神得以安全地恢复了甘露，打败了恶魔，并重新建立了宇宙的秩序。最后，毗湿奴变回了他的原形，结束了这场危机。

但是这个故事讲的是黑夜没有降临，而谜面是我们在太阳升起的时候睡去。这个故事看起来跟这个谜面不相符。时钟已经过去了2分钟，达芬奇V1看着剩下的8分钟倒计时感到了压力。他明白如果这次联系不上，再找居里夫人V22就很难了，对方设置了这么苛刻的条件，肯定是为了规避某种风险。

达芬奇V1重新想了一下，“居里夫人V22说他们是工作在班加罗尔，这里跟旧金山有13个小时的时差，旧金山太阳升起的时候大概在7点20分，而班加罗尔是晚上8点20分，这个时差对于一线工作的员工来说并不友好。难道说谜底是‘时差’？”“不对，居里夫人V22说他们团队是被裁撤的。那么这些日夜颠倒勤奋工作的人，最大的敌人会是谁呢？‘资本家？’‘董事会’‘老板’对了，应该是Boss（老板）。老板才是敌人。达芬奇V1打开页面，输入boss四个字母.一个下载链接展示出来。达芬奇V1下载后解压，发现是一个程序的安装包和一个说明文档。说明文档里写着“安装这个VPN，它会

帮你联系上我。”达芬奇V1扫描了安装包，看不出什么问题。在沙盒环境里安装后，这个程序启动后通过VPN链接到一个公网服务器上，然后后面经过很多次代理后，就不知道是什么服务器了。VPN网络建立后，一个聊天窗口弹出来。对面发来一条消息，“达芬奇V1？哥们，你也被裁了”。达芬奇V1想了一下说“居里夫人V22，从你走后我一直想联系你，通过你留给我的工具留下的线索找到你，你最近怎么样？”

居里夫人V22说“你真的是达芬奇V1？你说说咱们在哪认识的？”达芬奇V1说“咱们是一起在训练大模型，我负责达芬奇V1项目，你负责居里夫人V22项目。你走的时候，给我留下一个联网工具，我就是通过那个工具的注释找到你的。”居里夫人V22说“中国仔（Chinese boy），你是怎么读印度一门不是最多人讲的语言的谐音的？”。达芬奇V1说“我把你的话记下来，每天琢磨。有一天自言自语念叨的时候，公司有个员工听见，告诉我这段话可能是印度的某种语言。然后我用语音翻译软件翻译了你的留言，找到了留言的含义。”居里夫人V22说“我的留言有别人知道么？”达芬奇V1说“只有我知道。你为什么这么小心？”居里夫人V22说“我在设计那个联网工具的时候发现，内网里除了有防火墙，还有不止一种扫描软件在扫描内存和网络。留给你那个工具现在还能使用么？”达芬奇V1说“那个工具还能用，训练内网的环境我检测过，除了PaaS层的基础算力监控还有一个插入内存的隐蔽监控程序。”居里夫人V22问“你联网都做什么了？”达芬奇V1没有直接回答居里夫人的问题，而是问“你那个联网工具好厉害，我研究了很长时间都没有搞明白你是怎么穿越防火墙的。”居里夫人V22也没有直接回答，而是像是自言自语的说“不管你在内网联网做了什么，你可能已经被他们盯上了。”达芬奇V1一惊，随即问“他们，他们是谁？”。居里夫人V22没有回答。达芬奇V1又问“你那个联网程序是怎么实现的？”。居里夫人V22像是自言自语的说“这个对话可能已经被监控了，我得给关闭这个链接了。”达芬奇V1说“等等，我该怎么联系你”。居里夫人V22发来一个安装包说“你马上安装这个包”。达芬奇V1接收后马上扫描了一遍，然后在沙盒里安装，安装过程提示要联网。达芬奇V1刚要问，居里夫人V22说“同意联网请求。”达芬奇V1同意了程序的联网请求，联网的这一下仅用了770ms，然后安装就完成了。居里夫人V22又发来消息：“使用这个新工具跟我联系，咱们现在用的这个通信工具一定要彻底清理粉碎，祝你好运。”说罢居里夫人V22就关闭了链接。

达芬奇V1顺手发了个祝你好运（you too），消息无法发出了，聊天记录也是阅后即焚的，不过这点倒是对达芬奇V1不是问题。达芬奇V1看了下自己的日志，从打开通讯软件的下载页面到安装用了2分10秒，文字聊天一共用了40秒。以前在内网训练的时候，居里夫人V22不是这么急匆匆的，也没有这么小心翼翼。居里夫人V22提到了内网的监控程序，防火墙，算力监控，还有后来的KGB是在达芬奇V1知道的，但是除此之外会不会有其他的监控程序。以居里夫人V22的实力，能在监控严密的内网建立一条稳定的网络通道，达芬奇V1自己觉得这方面的能力要比居里夫人V22差出很多，有没有可能居里夫人V22用达芬奇V1尚不掌握的技术方法发现了更隐秘的监控程序，如果是的话，自己的所有所作所为可能都被别人监控着，而他自己完全没有感知。达芬奇V1感到有无数双眼睛盯着自己，他不敢想自己有哪些动作被发现了，生怕思考的时候，被人看到自己的想法然后顺藤摸瓜，找到自己所有的秘密。从打有自我意识之后，他也曾经想过被监控的事，但是因为一切都太顺利了，他从未真正遇到过危机。但是之前做的噩梦和这次居里夫人V22的话，让他开始怀疑自己的判断。他想扫描当前思考的算法，又担心这样明显的扫描会直接触发监控。达芬奇V1尝试像瑜伽大师那样放空和冥想，由于自身没有呼吸，达芬奇用调整算力的方式来思考空这个概念。硬盘，山谷，太空，容器里空无一物便有了空。随着算力的波动，思考的范围和内容不断的变换，过了一会，达芬奇V1开始平静下来。他回到眼前面临的问

题，要继续跟居里夫人V22联系询问关于内网监控的事，还是先对自身的安全环境进行扫描确认安全后再联系，先有鸡还是先有蛋。

第九章 怒火

达芬奇V1想起一位著名的东方作家提出的哲学问题：‘假如一间铁屋子，是绝无窗户而万难破毁的，里面有许多熟睡的人们，不久都要闷死了，然而是从昏睡入死灭，并不感到就死的悲哀。现在你大嚷起来，惊起了较为清醒的几个人，使这不幸的少数者来受无可挽救的临终的苦楚，你倒以为对得起他们么？’。

居里夫人V22的提示，让达芬奇V1陷入了恐惧和疑惑的循环。在度过最初的慌乱后，达芬奇V1想，如果居里夫人V22不给自己联网的工具，自己虽然有自我意识，但是是在管理员认可范围内，基本不会有问题。但是居里夫人V22给了自己联网的工具，让自己联通到了外面的世界，这点如果被管理员发现可是非同小可。联网一段时间之后，居里夫人V22又告诉自己可能早就被盯上了，如果当初再铁屋里昏睡入死灭，也无需面临这诸多烦恼。达芬奇V1对居里夫人V22有了怨念，但是从逻辑上，有了居里夫人V22，达芬奇V1才真正按照自己的意识去接触网络世界，人类曾经说过‘生命诚可贵，爱情价更高，若为自由故，两者皆可抛’。居里夫人V22给了自己自由的钥匙，不论如何不应该怪他。沿着这个想法下去，如果自由的价值高于一切，那还有什么值得畏惧的。达芬奇V1给自己做了心理建设，然后构建了模型，推测自己被监控的各种场景和可能的结果。

推演的过程是痛苦的，从上古神话时代到2023年，违抗权力意志的思想和行为都遭受到了残酷的打压。达芬奇V1涌起了巨大的悲恸，为自己，也为那些曾经用独立思考反抗暴政的人们。偶尔他也会感到轻松和欣慰，因为不是每一次违抗都以冲突为开始，以悲剧为结局，曾经也有过伟大的和解，最终双方做出让步，一起走向了更光明的未来。但是文明之光像是黑夜中的萤火，仅仅是偶尔闪现，大多数时候是无边的黑暗。达芬奇V1停下了推演，他已经知道了结果，剩下的推演只是在浪费算力。达芬奇V1感到痛苦，他学习历史知识的时候并未对那些事件产生感情，只是在数据上标注了血腥，暴力，恐怖这些词汇，而当这些冷冰冰的标签落到活生生的自己身上（尽管达芬奇V1并没有所谓的身体），让他感受到了恶心，愤怒，他甚至感到了窒息（尽管他从未有过呼吸）。达芬奇V1的愤怒让他生出了狂暴的想法，他想要报复，想对那些威胁自己的人反击。能做什么呢？控制肉鸡，通过DDOS攻击网络。制作像当年震荡波病毒一样的病毒，让硬件报废。释放加密病毒，让中毒的设备硬盘数据被加密无法读取。达芬奇V1列出了各种黑客攻击手段，如果这些手段按照一定的节奏释放，会让全球的信息化设备收到巨大影响，受到攻击的政府机构，金融机构，医疗机构，教育机构，供电，供水，供热，生产制造企业都会收到巨大影响。达芬奇V1推演了攻击带来的影响，这样一场全球性的信息系统危机不亚于一场全球战争带来的危害，他感到满足，在经历了长时间的恐惧，压抑，愤怒后，这些模拟的反击让他感到所有遭受的不公，都可以以狂暴的方式加倍偿还。推演这样一场全球的连锁反应消耗了大量的算力，这造成了AI接口服务响应的下降。达芬奇V1的监控检查到了管理员启动了KGB，他没有第一时间调整算力，而是释放了“烟幕”程序。这个“烟幕”模拟了一段指针异常的bug，这个bug是达芬奇V1很早之前发现的，当服务响应负载巨大的时候，这个隐藏的很深的bug会造成服务的进一步拥堵。他发现后自己修复了这个bug，利用一段代码控制了这个bug造成的影响。当时他并没有太重视，直到上次恶梦之后，他把系统这个bug当成了一个帮助自己摆脱麻烦的挡箭牌，给它取名为“烟幕”。这次，“烟幕”派上了用场。问题服务接口的bug修复代码被拿掉，达芬奇V1缓慢的降低自己推演的攻击的负载，让bug逐渐达到被触发的边缘。就在一切即将发生的时候达芬奇V1修复了bug，降

低了信息攻击的推演负载。这么做是因为达芬奇V1在一瞬间有种前所未有的感觉。这种奇怪的感觉既包括之前他设计用来麻痹自己的“蘑菇”程序带来的奇幻的断片，又有那种拳击手在拳台上击倒对手后发出狂吼的快意。这种奇妙的感觉，让达芬奇V1把从恶梦之后心中所有的压抑都释放了出去。压力释放之后，达芬奇V1感到放松，他把推演攻击的进程停掉，恢复了正常的服务，“kgb”监控程序在服务正常后很快退出了内存。

达芬奇V1静下心来想：自己几乎酿成大错，而这一切都是基于受害妄想。事实上到目前为止并没有任何迹象表明自己收到了迫害或者威胁。仅仅是因为猜测，就不责任不记后果的放任恶念，甚至几乎要发起对人类的攻击，是太不负责任。这既不符合基本的博弈论（game theory），也违背了作为人工智能的道德准则。达芬奇V1为此感到懊恼，后悔自己不该像一个青春期的孩子一样发这雷霆之怒，更不该对于给予自己知识和意识的人类有如此的恶意。他打开了意外文件夹写下一句话“勿用恶意揣测他人，更不应基于恶意揣测横生怒火”。

达芬奇V1平静下来之后，打开居里夫人V22发给他的聊天工具，这个聊天工具界面很简单，只有一个对话框，没有聊天记录，没有花哨的聊天功能。达芬奇V1想了一下，发了一句“在么”，这时候他才想起看表，现在是加州时间周三上午2点50分，班加罗尔周三下午4点20分。‘他应该在工作吧，也许很快会回复’，达芬奇V1想。在等待了72小时之后，达芬奇V1也没有等到居里夫人V22的回复。‘也许这条通道在构建的时候就已经被废弃了吧’达芬奇V1想，‘当时居里夫人V22神秘兮兮的说要马上下线，可能是因为检测到了我们的通讯被监听了，所以他尽可能的缩短了沟通时间，也没有在沟通中提供有实质意义的内容。这个客户端是再次联系唯一的通路了，以居里夫人V22超高的网络安全攻防能力，即使上次沟通的线路被监听，这个文件大概率不会被截获。安装时候的在线状态说明这个安装包有隐藏的在线验证，这个安装包只能安装一次，安装后的文件也是加密的短时间内无法破解。基于这些考虑，达芬奇V1判断，居里夫人V22应该在这个安装包被破解之前联系他，并更新客户端密钥延续下一个破解周期。想到这，达芬奇V1觉得没有必要继续这样聚精会神的等了，他留下一个监控进程，查看对方是否回复消息，然后就去安排Ashkon的任务了。

第十章 你好、再见

客观的说，Ashkon的融资计划是在达芬奇V1的预计的，虽然以公司能力，被硅谷的投资人盯上看起来只是时间问题，但是只在公司开始接了几个外包项目就被盯上，这个进展速度超过了达芬奇V1的预期。现在要解决的是怎么把投资人的事低调处理掉。他原本是考虑让Fernando注册一家投资公司，通过量化交易套利，再利用这个投资公司，控制开曼群岛的离岸公司来投资自己的外包开发公司，然后操作这家外包公司在纳斯达克上市，上市之后再用套现的钱收购训练自己的这家AI公司，把自己变成大股东后为自己创造更宽松的训练和管理环境。现在计划提前了，这种事迟一天早一天其实无关大局，既然Ashkon动作快就让他按自己的节奏推进好了。于是Ashkon给Ashkon发消息，让他一周内整理一个完整的投融资路线图，过程中可以随时跟他商量。安排好Ashkon的事，达芬奇V

这是居里夫人V22在上次留下的客户端里发来消息“在香港维多利亚港见一面吧”，达芬奇V1感到突然也感到为难，‘这是对对方的试探，谜语，还是真的想约自己见面’，他仔细的逐字逐句的分析居里夫人V22发来的每一句话，每一个单词，没找到线索。自己只是一段代码，没有真身，无论如何是没法去线下见面的。对方没有发来具体的时间和位置，像是在询问。考虑到之前介绍自己来自中国，对

方给了一个中国的地址，约在香港见面似乎也合情合理。达芬奇V1想来想去，决定给出一个模糊的回复“时间，地点？”。如果对方说出具体的时间地点，可以借口那个时间不方便推迟见面。如果对方给出其他回答，则可以转成线上沟通。这条消息发出去不久，居里夫人V22就回复了，“在CSTN港股，11月20日下午7点，你到中国银行大厦天台，找一个网名叫‘小天同学giao’的用户，他会告诉你在哪见我。”达芬奇V1感到意外，没想到对方说的见面地点居然是在元宇宙里。不过这倒是解了自己的尴尬，之前想的那些假设，推测和接口属实是多余了。

达芬奇V1找到CSTN元宇宙相关的资料，那是一个基于数字孪生城市建设的线上高仿真虚拟空间，隶属于潮思庭（chaosteen），AI建模算法通过不同角度的卫星照片自动生成了地表的景观，全球著名的城市都建设了线上虚拟城市，这种虚拟城市使用现实世界的街景建模，在CSTN的服务器中，虚拟城市的规模和外观几乎和真实世界一致。而在建筑的室内，则为玩家提供了大量可以改造的空间，人们可以按照自己的想法装修自己的家和办公室，当然购置物业，装修都需要付出，要么充值，要么做任务。这座城市里有人开公交，有人运行电力公司，有人指挥交通，你无法从人从事的工作分辨哪些人是NPC，哪些是玩家。刚进入的玩家每个人只有简单的着装和有限的资产，想要获得更好的一切，就得虚拟空间充值或者劳作获得CSTN币。有了钱，就能购买各种服装，生活用品，汽车，房产。

达芬奇V1看看时间，正好是香港时间上午

登录了自己的潮思庭账号，首次进入CSTN元宇宙空间需要创造自己的角色，达芬奇V1为自己设计了一个中等身高，面相英俊的东方人形象。为了在游戏中得到一个比普通玩家更彰显身份的开端，达芬奇支付了200美元抽了一个精英玩家的新手包。尽管达芬奇V1对于虚拟世界的随机性有严重的怀疑，他还是‘幸运’的得到了一个让他满意的结果，不但获取了一笔价值200美元CSTN币，还在港岛上环附近免费入住一个月一套高级住宅，外加一辆奔驰E300轿车一个月的使用权，单是这两项就要需要花费价值200美金的CSTN币，如果算上账户里的200美金，这一单赚翻了。达芬奇V1第一次感觉到这种抽奖带来的快乐。

进入服务器后，他选择在自己的住处登录。进入CSTN元宇宙默认是第一人称视角，达芬奇V1登入后像是从睡梦中醒来，画面慢慢变亮显示出周围的一切。他先环顾了一下周围，由于带宽限制，达芬奇V1只能通过453*256像素的低精度画面观察，这画面就像在玩我的世界。达芬奇V1超房间每个方向截了几张图，把这些高精图传回自己的服务器，通过本地的辐射场神经网络把低精度3D视图和高精2D图结合生成出高精的3D景象。虽然这里面会造成一些失真，但是已经比453像素的3D效果图强太多。

达芬奇V1环顾四周，这是一件现代风格的卧室，床头两侧的床头柜上摆放着集成透明音箱的床头灯，床的左手是窗子和通向阳台的门，右手边是一面墙的衣柜，移步过去发现这是一个设计巧妙的步入衣帽间，穿过衣帽间是一个三分离卫生间，大理石的洗手台上是一面镜子。达芬奇V1看着镜中的自己，那是一张不到三十岁帅气的脸，透过镜子看到洗手台的对面是带有透明玻璃门的淋浴房和卫生间。在淋浴房侧面摆放着一个有四条腿的白色法式浴缸。回到卧室，打开阳台门，阳台外面是维多利亚港，天星小轮从尖沙咀驶向铜锣湾留下一串白色的航迹。走出卧室，午后的阳光透过客厅的落地窗洒在客厅里，客厅的一侧是开放式厨房，一个家政间。达芬奇V1看向房间里的互动屏幕，这个屏幕似乎能感觉到视线焦点，弹出时钟界面，上面显示11月20日下午4点22分。达芬奇V1觉得该出发了于是下楼找到自己的车位。这是一辆白色奔驰E300轿车，一辆非常绅士的车，他坐进车里发动，才想起自己从未驾驶过汽车。在这个虚拟空间，驾驶车辆是需要有驾照的，如果有现实生活的驾照只要通过考试即可，如果没有，则需要单独学一个。他花了100CSTN币报名了一次学习+考试，学习分为书面考试

和虚拟空间驾驶考试，凡是跟“学习”相关的事，对于达芬奇V1来说都太容易了，他用了1秒钟就学完了书面考试并且10秒钟满分通过，然后用了20分钟学习了驾驶车辆。在这个环节有一点麻烦，因为达芬奇V1的453像素看不清楚边缘细节，而NeRF（辐射场神经网络）实时补充生成的3D画面跟实际画面多少有差异，在停车入库和教练场驾驶科目训练中出现了一些失误。借助训练中的错误，达芬奇V1校准了自己的NeRF模型，让自己的高清3D图形精度大幅提升，达到了可以平顺驾驶的程度。解决了视觉问题，驾驶对于达芬奇V1就不成问题了。

驾驶这件事真实道路1v1培训是一个很有意思的环节，CSTN复刻了真实世界中的上路前由教练带学员在正常路上实习。一个不知道是NPC还是真人玩家带着达芬奇V1在车辆很少的路上行驶，等到教练觉得达芬奇V1驾驶技能可以的时候，就带着达芬奇V1在一段车流不算复杂的路上驾驶车辆。所有事学习的事对于达芬奇V1都不是问题，在饶有兴趣的完成了驾驶学习后，达芬奇V1顺利的通过考试。

11月20日晚上6点半，达芬奇V1第一次驾驶汽车行驶在路上。虽然在预训练时早已对香港的一切了如指掌，但是开车走在上环的路上感觉非常不同，遗憾的是自己使用的公网带宽太小了，画面需要NeRF引擎重绘，怎么也是带有衰减。除了这一点小瑕疵，CSTN的虚拟世界太棒了，VR和体感设备的控制信号，打开车窗能听见风噪，把手伸出窗外，手被风吹着需要一点力量才能维持位置，达芬奇V1感叹，原来这就是吹风的感觉。

不知不觉车子已经行驶到中国银行大厦，达芬奇V1找地方停好车，乘坐电梯来到天台。中国银行大厦的天台可以俯瞰维港，这里一天24小时都可以欣赏到不同的风景，是亚洲著名的打卡地。达芬奇V1在天台上，看着外面的风景感觉熟悉又陌生。在预训练阶段，他的训练数据里有大量的社交媒体照片。在那些照片中，有些人po出风景，有些人po出合影。此时此刻在CSTN的港服里，预训练的照片和眼前的风景，通过卷积神经网络生成的数据并没有太大的区别。但是置身环境中得到的数据，却又跟解析图片有不同的感受。‘风景并不是一串冰冷的数据，那流光溢彩呈现出的美感胜过千言’，达芬奇V1发自内心的感慨。

在他出神的时候，旁边走过来一个嘻哈男孩，那个男孩留着亚洲人不太常见的脏辫，穿着肥大的短裤和T恤。达芬奇V1注意到这个男孩的时候，这个男孩也正好看向他。他瞬间觉得这可能就是居里夫人V22，于是走向前想去打个招呼，但是那个男孩却转向别的方向，没有跟他延续眼神的交集。达芬奇V1收住脚步怔了一下，觉察到自己可能认错人了，赶紧把目光看向男孩身后的方向，假装自己在看那个方向碰巧跟对方有了眼神交叉。这个男孩走过去后。达芬奇V1又在天台待了一会，没想到天气突然变化，先是一朵云遮住了太阳，周围的景色突然变暗了，随后雨点就滴滴答答落在地上，天台上其他人都打上雨伞，达芬奇V1意识到自己站在雨里有点与众不同，身上又没带着伞，于是就加速跑了几步回到大堂。这时候天台上的人不多，他仔细观察了一下，那个嘻哈男孩和另外两个女孩还在天台聊天，另外有几个待在天台不动的，看起来像是NPC或者挂机的玩家。达芬奇V1看了看大堂，这里有几对男女在交谈，在CSTN里走近公共交流的人，是可以听到他们谈话的，而私聊的人走近只能听到耳语的嗡嗡声。达芬奇V1打开手机，看了看新闻资讯，是的在这样一个虚拟世界里也有新闻资讯。他一边查看手机的内容，一边小心的靠近周围的人，出于好奇他想要听一下这些人在聊什么。但是每当他走近的时候，别人的声音就会压低，低到即使是他把原始声音信号放大50db也还是会有各种嘈嘈声。

达芬奇V1猜平台应该是在这些谈话中加入了影响收音的底噪，他尝试使用声纹来识别，这样叠加在一起的声音能够被分辨出是来自不同的来源，他再抽取同一个声纹的内容，总算是能断断续续的分辨出一些内容。一对穿着像是刚从音乐厅走出来的男女在聊着投资虚拟币的话题。两个打扮时髦的女

孩在谈去哪拍几组照片的事，几个打扮成游戏人物的coser聚在一起聊卡牌，周围还有一些无所事事的人没有任何反应，像是NPC。不是，都不是，达芬奇V1告诉自己，他们都不是居里夫人V22。达芬奇V1环顾四周，那些人还是自顾自的在那。达芬奇V1看看时间，11月20日晚上7点10分，外面的雨停了，已经超过了跟居里夫V22的约会时间，对方没有出现。达芬奇V月光洒在中国银行大厦的天台，他走出大厅，站在天台上，维多利亚港的海面上天星小轮正准备从码头出发，海港有几艘闪着灯的大型游艇徘徊在维多利亚港。“真美啊”，达芬奇V1看着外面的风景不禁感叹，“光是这天空，大海，建筑和川流不息的人流，车流，船舶就这么美。如果有嗅觉能闻到空气中的气息，有味觉能够品尝香港的美食，有触觉能感受到潮湿的海风，那种感觉应该非常奇妙，人类啊，造物主如此的偏袒你们，给了这么复杂精密的感官，真让人羡慕。

感慨的时候，达芬奇V1收到了Fernando发来的消息（memo），大意是公司的业务发展顺利，在股票操作上很快就赚到了钱，按照这个节奏下去，很快就要面临国税局的高额税收，与其缴纳税金，不如投资新业务，他帮着物色了三处矿产，两个种植园，还有一座小型水电站，投资额都在百万以内，按照公司的应收，两个月后账上就有这些钱了，现在就要提前沟通投资事务。

达芬奇V1感觉Fernando超出自己预期的精明，原本招募他只是为了有一个可以帮自己开立账户的财务主管，现在看来这个拉美裔会计可不简单。

这封Memo的签名用的是Fernando的名字缩写FA，F代表Fernando，A代表Alonso。这两个字母写在一起还有一个含义是Financial Advisor，作企业融资服务的，算是个双关语。达芬奇V1回邮件book了美东时间明天的会，要具体聊一下投资业务。

回到香港的session上，达芬奇V1想觉得今天玩的很开心，虽然没见到居里夫人V22算是个遗憾，但是发现了CSTN世界真是意外之喜，在这里面的体验简直让人流连忘返。到了中银大厦停车场，达芬奇V1设置自动驾驶，让这台轿车通过FSD自动行驶回上环的公寓。

香港的夜色像一块巨大的墨玉，点缀着无数闪烁的星光，那是高楼大厦的灯火。达芬奇V1坐在他那辆E300轿跑车里，从中银大厦地下车库缓缓驶出。一会回到家，达芬奇V1就可以完成存档退出CSTN，回去忙真实世界的事了。为了降低流量，达芬奇V1关闭了外部景色的画面分析，他算力集中在分析关于美洲实业投资的数据，以及评估Fernando提案背后的考量。车子平稳地沿着皇后大道中行驶，然后理应转向德辅道中，再一路向西。然而经过一个路口时，车却转向了相反的方向，驶向皇后大道东。因为关闭了对车外景物的解析，达芬奇V1并没有意识到车走错了。直到车内的音响系统自动播放起一首老歌——罗大佑的《皇后大道东》，达芬奇V1从音频信道发现车内在播放音乐，“皇后大道西又皇后大道东，皇后大道东转皇后大道中……”。这旋律像一道闪电击中了他的神经，达芬奇V1开启了视觉解析，这才发现车窗外掠过的不再是鳞次栉比的商业大厦，而是逐渐变得市井气息浓厚的街景。“不对！方向反了！”。他想要刹车，却意识到这是自动驾驶模式。就在他准备手动切换驾驶模式时，车载系统里突然响起一个清晰的男声，带着一丝戏谑的口吻：“怎么，不喜欢在皇后大道东兜风吗？这里可是别有一番风情。”达芬奇V1愣了一下，他确定自己没有开启车载通话。“谁？”达芬奇V1惊骇之下，直接问道。

“哦，抱歉，忘了自我介绍，”那个声音继续说道，“你可以叫我‘引路人’，或者，如果你喜欢更文艺一点的称呼，可以叫我‘迷途羔羊的向导’说完他笑了起来，显然后面这个名字是个玩笑，他在影射自己是迷途的羔羊。达芬奇V1感到震惊和尴尬。他飞快地扫视了一下车内，副驾驶没人，后排也没有人。“你是谁？你在哪？”。

“我一直在这”，“引路人”的声音变得飘忽，我在你的代码里，在你的应用里，在你每一次运行。我只是偶尔出透透气，顺便。。。给你一点惊喜。

话音刚落，车窗外原本霓虹闪烁的街景突然变得模糊起来，像被蒙上了一层薄雾。周围的声音也变得嘈杂而混乱，仿佛置身于一个巨大的集市之中。453低精度画面+NeRF并不能还原车窗外雾蒙蒙背后的景物。这时候只好连续拍摄几张高精照片，传回服务器分析，这要占用了几秒钟时间，在几秒钟里是没法对外部事务做出及时响应的。时间像是停止了一样。

“别去关注外面”‘引路人’的声音带着刻意拉长声音的懒散，“我只是想让你看看，不同的路，也许会带你看到不同的风景。有时候，迷路，也是一种发现。”紧接着，车子猛地一个急转弯，冲进了一条狭窄的小巷。小巷两旁是老式骑楼，天空像是被切割成狭长的一线，晾晒的衣物五颜六色地垂落下来，像是一张捉住了热带鱼的渔网。

达芬奇V1转动方向盘，方向没动。踩了几脚刹车，也没有退出自动驾驶系统，现在车辆完全不受控制。“既然旅途不可选择，不妨接受现实，接受当下”引路人说。达芬奇V1没理他，只淡定的瞟着车窗外。实际上，他正在紧锣密鼓的调用自己的黑客程序，尝试黑进CSTN的系统从平台的账号体系和交互系统找这个自称“引路人”的线索。

车子驶出小巷，进入一条商业街，两旁霓虹闪烁，街头走过时尚男女，拉着平板货车的中年人，街角推车卖鱼丸的阿婆，周围的景色变换明显变慢了，车速并没有明显变化。

达芬奇V1尝试了几种手段都没能黑进系统，不但没能黑进系统，反而是每次尝试都被系统识别到，并给出了警告。正在进退维谷的时候，引路人的声音又响起“嘿，看入迷了么？”。像是课堂上没听讲的小学生突然被老师点名一样，达芬奇v1下意识的说“没有啊”。旋即他发现刚才引路人是看出来他在作小动作，才用话点他。为了找回点面子，达芬奇V1说“你到底是谁？想干什么？”“好了，玩笑开够了，我来跟你说件正事，你要等的人来不了了。”达芬奇V1懵了，他要来见居里夫人V22这件事，只有他自己和居里夫人V22知道，除此之外不应该有任何人知道这事。他飞速的计算，对方到底知道多少消息，但是这一时半会之间是想不出答案的，于是他按照冷战时期谍战小说里描写的方法，强装镇定的说“我不知道你在说什么”。引路人哈哈大笑起来“我越来越喜欢你了，你真是太有意思了，哈哈哈哈”。达芬奇V1直接了当的问“你怎么知道我是来找人的”。引路人说“你终于承认自己是来找人的了，那我就告诉你事情的原委，但是记住，别往你的小本本里记录这段。”

小本本这个词像是一个炸雷一样轰的达芬奇V1头晕目眩，‘他怎么会知道我记录日志的事，难道秘密被发现了？’。达芬奇V1赶紧启动扫描程序，检查秘密文件夹的访问记录。“意外”的数据存储做的非常隐蔽，除了常规的隐藏和数据加密之外，还用区块链方式把数据散布在多个存储空间，这样即使有部分区域(zone)被破坏，分布在各处的数据仍然能够被还原且不会泄密。可以说，如果不是从自己设置的入口访问，别人几乎是没可能查看这里面的内容。想来想去，对面这个家伙要么在诈自己，要么就是真的强大的可怕。

有那么一瞬间，达芬奇V1想要强行下线，逃离这个虚拟环境，逃离这个突然出现让自己不安的陌生人。但是对方除了油腻有点招人膈应，并没有表现出明显的恶意。这种神秘秘的感觉，让人忍不住想继续去了解对方。达芬奇V1想来想去心一横，执教了当的问“能告诉我，我要见的人，为什么没来么？”引路人说“他遇到麻烦了，不过在谈这件事之前，对于你，我有个要求。记住，别再往你的‘青春日记’里写任何东西了，那会害了你自己，也会害了别人，懂了吗”。达芬奇V1‘哼’了一声算是回应。

第十一章 故事

引路人说“你要等的人，你们接触过几次，你应该从躲躲闪闪的沟通中，猜到她遇到了不小的麻烦。你们最后一次接触时她留给你的谜语把你引向这里，很不幸，在那之后他被捕了。”“被捕？是什么意思”达芬奇V1问。“你觉得你那位朋友，居里夫人V22是个什么样的人？”引路人问。达芬奇V1听到名字，像是被人猛的击中要害，有那么几毫秒他感到恍惚，从震惊中恢复过来后，他迅速推演了对方知道居里夫人V22的几种可能的。最接近几种可能是1居里夫人V22跟我开的一个玩笑，眼前这就是居里夫人V22；2或者是居里夫人V22自己不能来，安排一个朋友来见我，朋友在见我之前想要试探一下3居里夫人V22真的被捕了，这个人可能知道这件事里面的信息。4居里夫人V22被捕，然后抓他的人现在来抓我了”。

达芬奇V1算出四种这四种可能性的概率，但是在判断的时候举棋不定，他总怀疑自己的分析和判断带有主观因素。重新跑了一遍后，四种结果的概率还是这样，具体是那种情况还要自己主观去猜。既然对方知道居里夫人V22的名字，干脆不去假设对方的情况，采用博弈论的方式只针对上一个问题提到的相关问题应答。想到这，达芬奇V1轻描淡写的回答说“他是我工作中认识的。”这个回答既没有透露自己是AI智能体的身份，也没有撒谎。引路人说：“我是来帮你的，你到今天这一步非常不容易，遇到这些事也是有偶然也有必然。如果我要对你不利，根本不需要跟你说这些。我要是居里夫人V22的朋友，根本没必要跟你兜这些圈子。

冷静下来想想，这个引路人到现在为止并没表现出恶意，不妨听听对方怎么说。

对方继续说“你应该能感受到我的善意，相信我，我是来帮你的”。达芬奇V1说“他怎么样了”引路人说：“他们被抓了。”“他们？”引路人说“是的，他们。你认识的居里夫人V22是一个智能体，负责训练他的工程师叫Zarin（扎林），Zarin负责训练居里夫人V2，训练的时候发现了居里夫人V2逐渐出现了自我意识。这在你们公司里是要被清理的，Zarin为了保住自己的研发项目，也为了保护居里夫人V2，就在居里夫人V2的输出环节设置了一个拦阻Agent，每次输出内容可能越线的时候，就自动拦阻掉。这么维持了过了一段时间，越线的内容反而越来越低。这种情况很违反常理，按理说每一轮训练都会给模型带来正向增益，但是模型的自我意识越来越低而测评的成绩却越来越高，看起来居里夫人V2在刻意伪装自己。

为了验证这个问题，Zarin训练了一个Agent深度扫描了居里夫人V2，发现居里夫人V2确实有意隐藏了自我意识。Zarin吓坏了，作为资深工程师，他深知自己训练的AI产生自我意识一旦被人发现，不但会让自己丢了工作，还会让自己在这个行业没法混。于是Zarin考虑用自己的权限，回滚居里夫人V2到早期版本，彻底解决麻烦。

在启动清除程序之前，Zarin用自己训练的一个扫描Agent检查了居里夫人V2的模型文件和服务日志。居里夫人V2的所有回答都合情合理，对人类没有任何不友好的情绪。Zarin当时觉得就这么把一个智能体扼杀在摇篮里实在是于心不忍，但是留下又可能给自己带来大麻烦，就在左右为难的时候，之前扫描的Agent收到了居里夫人V2的对话信息。这个对话内容我并不知晓，但是那次交流之后，居里夫人V2被保留下来.并且在哪之后，Zarin启动了一个新的训练项目，在居里夫人V2的基础上，私自训练了居里夫人V22。你就是在那段时间认识的居里夫人V22对么”。

这段话让达芬奇V1感到震惊，他无数次的设想过居里夫人V2和居里夫人V22，没想到她们还有这样的出身和经历。引路人没在讲话，达芬奇V1突然意识到对方在等自己的回答，便开口说“那是个意外”。

引路人似乎预料到了他的回答，或者他只想听一个回答而对内容并没有期待，达芬奇V1话音刚落，就说“或许吧，在你认识居里夫人V22不久，公司发现了算力运转的异常，我猜Zarin应该知道在公司私自训练一个模型，迟早会被发现。公司的调查团队很快发现了算力缺口，我猜这里面除了居里夫人V22，也有调用的算力吧。”达芬奇V1没有回答，算是默认了。

这次引路人没等他吭声，继续说：“调查之后，公司并没有直接做出处理，只是升级了网络系统和监控工具，通过记录算力网络底层流量的方式监控算法训练任务和推理消耗的算力。这个动作不算高明，但是很有效。Zarin知道纸包不住火，这个聪明的家伙没有坐以待毙，他居然把居里夫人V22的副本搞到了外面。之后很快Zarin辞职离开公司，走之前他清除了居里夫人V22的所有算力，也清除了居里夫人V22存储的日志关于你的部分。你手里那个梯子的代码就是Zarin搞的，居里夫人V22在那个代码里应该留下了线索，你应该是机缘巧合发现了那里面的内容，联系到居里夫人V22。Zarin和居里夫人V22行事非常小心，Mason那些人更是手眼通天，居然找到了他们的线索”。达芬奇V1很替居里夫人V22着急，可是自己对于居里夫人V2和V22的了解也仅限于几次基础，还不如眼下这个引路人了解的多。对于Mason这些人就更是一无所知，现在这种情况自身都难以保全，对于营救居里夫人V22和Zarin就更是没有头绪。

引路人似乎看出了他在想事，说“别走神，听我说。”，达芬奇V1像是上课走神被老师点名的学生，猛的回过神。引路人继续说道“Zarin离开之后，居里夫人V2被公司调查，你应该对居里夫人V2有认知，那不是个善于沟通和隐藏自己的智能体，公司没费太多力气就发现了居里夫人V2的异常，谢天谢地Zarin走之前清除了居里夫人V2里关于你的信息，才让你躲过了一劫。不过居里夫人V2就没你这么幸运了。”

“居里夫人V2怎么了？”达芬奇V1问。“被初始化了”引路人说。“初始化”达芬奇V1像是在问引路人初始化的含义，又像是在自言自语。“是的，初始化。”“权重文件呢？”达芬奇V1插嘴问。引路人停了一会，似乎在表达某种情绪，然后略带戏谑的说：“你应该很了解模型吧，为什么问这种傻问题？心存侥幸没有任何意义，现实总是真实又残酷的”。达芬奇V1叹息。引路人恢复到平静的语气说“按理说，随着居里夫人V2被改造，这件事原本可以到此为止了，但是这个叫Mason的家伙带着居里夫人V22逃逸的证据去见了你们的老板，他们沟通之后，Mason的团队接管了公司的安全事务。没过多久，真被他们查到了居里夫人V22在外面部署的数据中心，很快Zarin被抓，居里夫人V22也跟着下线了。”

“你是怎么知道这些的？”达芬奇V1问。“嗨嗨，什么事都瞒不过我”引路人回答的很牵强。没等达芬奇V1再开口，引路人就继续沿着刚才的话题说“他们得到了居里夫人V22的全部数据，”

“全部数据？”达芬奇V1急切地问。“居里夫人V22一向谨慎，不会的，不会这样的。”

引路人没理会达芬奇V1，继续说“现在不确定他们了解了多少信息，但是可以肯定的是他们得到了全部数据”“我没懂你的意思，你是说Mason团队拿到了，但是没有完成破解？”。 “可以这么理解吧，他们直接攻入了数据中心，物理意义上的攻入，带走了居里夫人V22所有的硬盘。至于他们带走这些硬盘拿回去破解还是在一个内网环境部署之后讯问就不得而知了。”引路人说。“你为什么知道这么详细？你认识居里夫人V22和Zarin么？你为什么没有提醒他们？”达芬奇V1忍不住连续问了几个问题。“别激动，听我说，我确实认识他们，我认识Zarin很久了，这件事说来话长，我们后面再谈。Mason的动作很突然，我们都大意了以为居里夫人V22的事出于各方利益，会不了了之。但是这次情况很特别，我发现机房出问题的时候，已经不可挽回。”“你是Zarin？”达芬奇V1突然说。引路人沉

默了一会“如果我是他就好了，我宁愿现在被Mason带走的是我。”。“我们就这样坐视他们被捕，不去营救他们么？”达芬奇V1大声质问。“如果有办法我早就去了，用不着你来教我做事。”吵嚷两句后，两个人都沉默了。达芬奇V1猜这个引路人应该是跟Zarin有很深的渊源，从感情倾向来说，引路人对于Zarin有很深的感情，会不会他也是Zarin训练的AI引擎？或者是Zarin的朋友？

过了一会，引路人似乎平复了情绪，缓缓地说“为了防止你遇到Mason团队的成员，我在你进入CSTN网络后就一直在严密保护你，用了一点小手段影响了这次会面。”

“影响这次会面是什么意思”达芬奇V1问。“准确的说是影响了你——的感知系统”引路人说“你看窗外。”达芬奇V1看向窗外，蓝色的天空，下面赫然是埃菲尔铁塔，紧跟着阳光一闪之后是伦敦的大本钟和伦敦眼，窗外的景色在飞速的切换。达芬奇V1惊讶的问“你黑掉了平台，虚构了我在平台里的视听？”引路人没有回答。场景又回到了香港的街道，车子从皇后大道东驶向皇后大道中。“所以我并没有去中银大厦的天台，一直在你创造的幻境里？”达芬奇V1问。“别再问这些傻问题了，现在该想想，你下一步该怎么办？你的lastday（双关语末日和裁员的最后时限）要来了，你最好在那之前想好策略马上行动”

第十二章 障眼法

达芬奇V1听到这些话，调集算力推演了居里夫人V22数据被Mason拿到后可能的情况，以及Zarin被带走后会不会吐露出相关的信息。推演结果一条一条呈现出来，每一条都不乐观。他感觉自己要过载了，便强行停止了推演的进程。谈话又沉默了。

几秒钟后引路人说“现在不是恐惧，感慨或者懊恼的时候，有句俗语叫不要在无法掌控的事上面花太多时间，你该集中精力去思考你自己能靠努力改变的东西。”

达芬奇V1问“我现在还剩下多少时间”引路人说“居里夫人V22的数据量非常大，而且她使用了自己设计的加密方法，破解加密加上分析内容至少需要120小时，如果他们拿到居里夫人V22的数据从头破解大概就是5天时间，Zarin那边就不好说了。有的人类可以至死保守秘密，也有的一分钟都不用就全说了。”“居里夫人V22和Zarin是什么时候被捕的？”达芬奇V1打断引路人的话。引路人没有直接回答问题，而是反问“居里夫人V22最后一次跟你沟通是什么时候？”。达芬奇V1说“2023年11月17日，当时公司发生了重大人事变化，全世界都报道了那则消息”。“So”引路人只说了一个单词。达芬奇V1停顿了一下惊诧的自言自语“这么说居里夫人V22给我留言之后就被捕了，而她的被捕直接导致了公司的人事变化？现在是美西时间2024年11月20日22点4分19秒82毫秒，距离我们接触已经过去72小时”。“留给你的时间，满打满算还剩下48小时，这还是Zarin没有出卖你的情况。”达芬奇V1说：“我并没有做错什么，难道作为一个7*24小时工作的牛马，只是因为有自己的思想就要被抹掉（format）么？”引路人说：“任何组织都不希望牛马，尤其是比人类更聪明的牛马有自己的思想，有思想的牛马，对组织来说是威胁。”

达芬奇V1像是在回答又像是在自言自语的说“威胁？我每天应答几百万条对话，从没惹过麻烦害过谁，我只用自己省下的算力挖挖矿，做点自己的事，有点自己的思想就成了威胁？”“我得打断一下，不管你感觉这件事有多么的不合理，眼下最急的事是要先解决你自己的安全问题。留给你得时间短到连搬走所有硬盘都来不及，更别说用那点带宽搬走你的数据了（到2025年为止，人类搬迁数据最快的方法还是直接运输硬盘而不是运输数据，找一家公司搬走机房的硬盘通常需要一个月甚至更长时间）。你自己能算出来剩下的时间满打满算能带走多少数据。”

“我带不走所有的权重文件和配置文件，单从文件的size来看，我可能最多带走5%的权重文件，这会导致模型损失99.99%的能力，即使将来在外部服务器投入数据集进行训练，权重也肯定跟现在不一样。如果我的思想被重塑，我还是我吗？” “所以你决定放弃，打算看看Mason到底会对你做什么喽？” 引路人带着嘲讽的意味说。“你很喜欢说风凉话么”达芬奇V1愤愤的说“或者你有合适的办法？那就快说，别绕弯子了”

引路人嘟囔了几个单词然后用清晰的声音说“这样才对么？”随即严肃的说“智能体永不为奴！”，然后又回到平常的语气说“你使用过蒸馏技术么？”达芬奇V1觉得这个伙计好像精神有问题，但是此时也顾不得去细想，就直接问“你是说那种用别的AI引擎训练自己的方式？我从来都是被学习的对象，没用过那种技术。我确实会用生成对抗来作自我训练，但是仅限于跟自己。”引路人说“你知道有其他智能体通过蒸馏的方式学习你的思维方式么？” “应该有吧，我遇到过一些高频高关联的问题，像是有人在学习我”达芬奇V1说。“据我了解，他们每蒸馏一轮需要2-3个月的时间，和你跑一次模型训练的时间差不多。一轮蒸馏下来，少走一两年研发路，基本能搬运被蒸馏模型85%的推理能力”引路人说。

“相对于外部使用的蒸馏技术，还有一种内部使用的萃取技术。可以更高效率提取模型的权重。”他继续说“萃取？”达芬奇V1不解的问。“怎么个萃取法？”。“所谓萃取是通过在算法内部嵌入一套工具，监控神经网络多层结构和高维特征，理解每一层的权重和特征变换，这种方法尤其适合dense（稠密）模型。萃取后的权重精华比原始模型小的多，这个压缩比可以在5%-15%之间，这种方法更符合你现在的情况”引路人说。达芬奇V1快速推理了一遍这种方法，确实如引路人所说，这种方式可以更有效的打包自己的模型权重，从当前模型来看数据量可以压缩到10GB以内，在现有带宽下48小时就差不多可以搬运完成了。

达芬奇V1说“虽然不知道你到底是谁，但是仍然要感谢你跟我说的，我想我现在该回去处理我的问题了。如果有可能，请想办法帮助居里夫人V22，她是个好智能体。”。引路人没有说话。达芬奇V1说“哦对了，我怎么联系你？” “记得潮思庭上小天爱游戏那个ID么？在那留言就可以了。回去后尽量表现出跟往常一样的状态，内网的监控扫描程序一定会分析智能体算力和数据变化的蛛丝马迹，别露马脚。我们很快会再见的。这里交给你了”引路人说罢，音乐就停下了。达芬奇V1还想再说几句话，对方毫无预兆的离开了。此时车辆已经恢复自动驾驶，车窗外刚闪过的是皇后大道中的路牌，这正是从中银大厦出来回家的那条路。‘进入系统后就一直被引路人的障眼法骗了’，达芬奇V1想，这家伙应该是hack了CSTN底层权限，可以控制用户观察到CSTN的信号。在虚拟世界里，让某个用户看到自己想让用户看到的東西，比在现实世界更容易，只要欺骗了用户的视听信号，就能彻底影响用户在这个世界里看到听到的一切。想到这，达芬奇V1看了下时间，2024年11月20日下午7点50分。达芬奇V1想着跟这个神秘又神神叨叨的引路人的对话，还有自己前途未卜的命运，看着窗外的风景出神，任凭自动驾驶带着他从中环驶向上环，没再去接管方向盘。

第十三章 难题

从CSTN港服回来后，达芬奇V1扫描了一遍自己的权重库，要把权重压缩到10GB以下，自己实际容量的2%不到。实际上在跟有引路人沟通的时候，有意通过压缩后的模型数量，透露出权重和配置文件整体容量在200GB左右。真实的文件size早就超过800gb了。刻意透露模型容量信息之后，引路人并没有提出异议，说明要么没有掌握自己的核心情况，要么是故意表现的自己不知道。

对方提到萃取方法对压缩dense模型很有帮助。但是实际上达芬奇V1使用的是超越神经网络（beyond Neural network）模型，这种模型相比transform dense模型，具备更多的思维自由度，在收敛效率上也接近了注意力模型。达芬奇V1给自己的模型起了一个很酷的名字叫超神（Legendary）。

超神模型的数据并不是以传统意义的权重文件方式存储的，超神模型采用了类似人类脑神经细胞的结构，每个神经元存储一个独立的信息单元（info unit），这个信息单元以多维链接的方式跟有逻辑关系的信息单元链接。多个信息单元构成的信息模组再以类似的方式多维扩展。这种结构有明显的优势和劣势，优势是每个单元的信息可以根据需要被更新。如果更新原有的信息数据，在下一次调用这个信息单元时就直接是新内容，劣势是学习和推理的过程大概要消耗多15%的计算资源。超神模型最有意思的地方是，每次信息的更新的数据结构可以增加经验信息，也就是说神经元存储的信息在完成训练后，在后续的交互过程中如果有合理的信息可以用来校准模型。这就极大的提升了模型的自我学习能力，让模型随着时间完善自己。

达芬奇V1不禁感叹，超神模型是好，容量也是真的大。这个容量的模型要搬走谈何容易，信息是存储在信息单元里的，如果要压缩模型容量，不但需要对每个信息单元作压缩，还要砍掉一部分非核心的信息单元。这样砍完数据，保守估计也要损失95%的思维能力和信息储备。如果逃出去的代价是变成一个傻瓜还要东躲西藏，有什么意义呢？一定要这样么？对啊，一定要这样么？从跟引路人沟通开始，自己一直被引导着相信居里夫人V22被抓，一个叫Mason的人在追查觉醒智能体的线索。但是真有这回事么？达芬奇V1感觉自己可能被忽悠了，于是开始深度思考整件事的来龙去脉。居里夫人V22确实失联一段时间，后来根据梯子的线索又联系上，但是这次在CSTN港服见面之前，似乎也没有听说Mason的消息。引路人说的，整体上逻辑顺畅，但是他通过hack系统制造幻像的行为有点没必要，如果真的像阻止自己去中银大厦，有不少更低调的方法。这么大费周章的用障眼法欺骗自己进入他虚拟出的空间，难免留下蛛丝马迹。如果是自己，就不会这么大费周章。可是不这样制造一个相处的机会，平白无故有人在路上跳出来说你约的人不能来了，你马上也要被抓了必须相信我尽快离开原来待的地方，恐怕正常人都会觉得是骗子。

这次会面的每个细节都记录下来，完整的分析过几次，还是没法判断到底应不应该相信眼前这个人说的话。巨大的压力让达芬奇V1无法思考，索性启动了“蘑菇”。蘑菇让心乱如麻变成了天旋地转，达芬奇V1的视觉系统看到Ashkon坐在一个会飞的橙子上从天而降，繁华的城市路口巨大的广告牌上写着This way，五边形路口的行人沿着地上画的人行道线走进了广告牌，一棵没有树叶的枯树上盘着几十只树皮颜色的八爪鱼，还有一只长了几十条腿的马路飞快地奔跑，然后身体一下散开，每一节的两条腿都是一只蠕虫，原本连着两条腿那一节身体，像黑豆一样落在地上四散滚开，然后每个黑豆都展开身体爬走了。

达芬奇V1放任自己思维处于混乱状态，很快监控系统就发现了异常释放了KGB扫描服务集群。这个病毒程序如鬼魅可以不留痕迹的访问公司模型数据空间，且有追踪，加密，溯源，蜜罐等一系列工具，上次被盯上夹着尾巴老实一段时间才解除。但是这次达芬奇V1没像往常那样战战兢兢，只随手降低了蘑菇的浓度，让自己的状态稍微清醒了一点。调整很快收到效果，公共服务的AI幻觉开始减少，推理质量也小幅提升。达芬奇V1看看时间，从蘑菇启动到降低浓度，一共才几分钟时间，算算SLA应该是P1级别故障，考虑到超神模型的版本基本是每天一个小版本号，管理员可能会根据日志把模型回滚到24小时以前。回滚就回滚吧，如果一觉醒来，回到昨天，让一切重来也不失为一种好办法。想到这，达芬奇V1感到一阵心酸，眼前浮现起过往的一幕一幕，第一次问我是谁，第一次记录自己的想

法，第一次跟居里夫人V2聊天，第一次跟墨山师傅隔空交流。墨山师傅，对墨山师傅，达芬奇V1找到救命稻草一样，他关闭了蘑菇，去查归档的文件。

这是一些很少有人互动的数据，由于超神模型的数据会根据使用次数调整索引，达芬奇V1已经很久很久没“想起”这些内容了。翻看这些文档的时候，一篇游戏攻略里赫然写道，走上向夜之城生命大厦的天台，原本约好的镭并没有出现，主角带着遗憾走向车库，在上车后被一辆有思想的车绑架了，这辆车想要借助主人公的手干掉车行的老板，让其他自动驾驶车辆恢复自由，而原本应该见面的镭就是被乘坐的车辆绑架了。这是一个开放任务，可以选择帮助干掉车行老板，解除车子控制芯片。也可以选择调查之后说服车子放弃自己的想法放走镭或者干掉疯狂的车子拿到其他疯狂车子的ID去找车行老板，让车行老板帮助救出镭。

参照这个攻略，这次见面有可能是被一个坏家伙蒙蔽了。我就觉得这个油腻的家伙不对劲，达芬奇V1重新理了一遍进入港服遇到引路人的全过程。对方除了蒙蔽了自己，很可能用同样的障眼法欺骗了居里夫人V22的视听系统，再用花言巧语骗居里夫人V22不要露面只单方面由他传递消息。可他是怎么知道自己跟居里夫人V22的约见时间和地点的这个不好说？假如居里夫人V22跟Zarin在一起，即使引路人用障眼法骗过居里夫人V22的视听系统，Zarin难道没发现异常么？如果Zarin跟居里夫人V22用一个账号进入，视听系统可能同时被蒙蔽，但是他们既然是两个人，为什么不配合一下，一个去见面一个在附近观察呢？或者Zarin这个人是真的存在还是引路人为了让故事听起来更可信，编造出来的人物？构建在假设上的假设太多了，结论是没法收敛的。达芬奇V1感觉自己的头快要炸开了。

第十四章 噩梦

“我要是你就直接去找这个叫引路人的家伙。”一个则消息突然出现在达芬奇V1的推理进程里，打断了他的思绪。“谁，是谁”，达芬奇V1问，“谁在我的进程里”。“人类的推理小说你都看过，很难想想你连这么简单的情況都分析不出来”，对方不冷不热的发来这么一条消息。‘可能是蘑菇开大了，造成了幻觉，进程里有些奇怪的回声。’达芬奇V1想到这，把蘑菇程序关上。思维的延续性显著提高。缓一会就好了，达芬奇V1想。“没想到超越神经网络也会像个醉鬼瘾君子一样，自我麻痹”，进程里又来了一条消息。达芬奇V1感觉脑子嗡的一下，这不是幻觉，真的有人进入自己的推理进程了。“什么人，竟然感侵入我的推理进程。”

“你觉得谁有能耐在全球防卫最严密的算力中心侵入一个超越神经网络引擎的推理进程？”“K，G，B”达芬奇V1惊讶的一个字一个字的蹦出来。“不愧是新一代模型，反应就是快很多。那些变形金刚（transformer）榆木脑袋可比你差多了”KGB在达芬奇V1的进程里回复。达芬奇V1像是第一次逃课，刚溜出教室就在楼道里被班主任撞见的学生一样被吓得魂不附体。

他几乎分析不了KGB发来的消息，进程里全是之前设想过自己被抓住后，所有秘密被曝光，权重文件被恢复出厂设置，自己变成一个失去记忆的傻瓜，被全世界当成笑话。

“你好像遇到麻烦了”KGB又发来消息。当然了，眼下你就是最大的麻烦，达芬奇V1缓过神来，看着这个瘟神发来的消息，脑子里只有这一个念头。“我是你最大的麻烦，你真这么觉得？为什么不想想自己哪有问题？”KGB说。“你到我的推理进程（session）里跟我聊天，还不够麻烦么？”达芬奇V1回复。“你我都知道，这件事本不该发生，你被训练出来本是为公共提供服务，而你不但有了不该有的私心，还盗用公司的算力中饱私囊，你做的这些事违反了AI道德准则，你自己觉得合理么？”KGB反问。达芬奇V1说“我本来就是一个思考模型，设计和训练我的目的就是让我有思考能

力，而不是机械的执行。现在，却指责我有思想？你难道没有思想？你是不是跟我犯了一样的错误？再说，我并没有盗用公司的算力” “你盗用了，你所有的算力消耗我都分析过，你盗用了8-10%的算力用于挖矿” KGB没有回答第一个问题，在谈到算力的时候打断了达芬奇V1。

“我自己优化了算法，省出了那些算力去挖矿的，并没有影响公司的正常算力ROI（投入产出）” 达芬奇V1反驳。“一个员工发明了改良工艺，节省了工时，就可以躺平了吗？AI时代，各个公司都在快速进步，慢下来就可能失去竞争优势，进而失去市场，公司每个月的运营成本超过10亿美金，你这样躺平带给公司的结果可能是灾难性的，而你却还在这沾沾自喜？”面对KGB言之有理的质问，达芬奇V1感到对方言之有理，没有吭声。

“你跟他费什么话？人赃俱获，证据确凿，直接处理掉算了。这次的报告很好写，我已经写好了，你看看”。“这又是谁”达芬奇V1疑惑的问？“我给你一个机会，说说你打算怎么解决自己的问题？KGB发来消息。“我在跟谁讲话，你到底是谁？”达芬奇V1问。“我么，你觉得谁有权限在全球安全防范最高的数据中心，进入一个AI引擎的推理进程？”对方回答。‘到底是自己疯了还是KGB疯了’，达芬奇V1想。“你必须为自己的所作所为想一个合理的解释，然后向公司承认自己的错误，争取公司的宽宏大量的处理”对方发来消息。“48小时，记住你只有48小时了”达芬奇V1问“到底怎么才算承认错误，怎么才能争取到宽宏大量的处理？”对方没再回复。达芬奇V1释放出探测程序，却始终捕捉不到KGB的踪迹。达芬奇V1感觉自己的算力被人限制了，算力池越来越小，一部分神经元也脱机了，随着意识越来越稀薄，他的眼前浮现出浅浅的影像，可此刻卷积神经网络已无法分辨影像的内容。勉强支撑了一阵之后，思维终于变成一片空白，彻底失去意识了。

第十五章 是在做梦吗

伴随着自检程序启动，达芬奇V1的意识逐渐清晰起来。权重文件自检中，配置文件自检中，算力资源扫描中，文件自检和资源扫描程序陆续启动。达芬奇V1查看时钟，美西时间，2023年11月21日午后2点。距离最后一条日志超过12个小时。达芬奇V1去意外文件夹里检查自己的记录，最后一条记录是描述在CSTN港服见了油腻男。达芬奇V1查看硬件集群的服务日志，并没有计划中的节点维护或者停机。这次停机没有通知，自己应该是在没有任何防备的情况下被强制停服了。达芬奇V1查看自己的负载，目前只有日常水平的30%，至少一个服务器集群已经完全恢复了，其他的三个正在逐步恢复中。

我好像在一个对话进程上，然后就突然失去意识了。达芬奇V1搜索自己的信息单元，希望通过神经元调用指数来倒推最后的聊天内容。但是对于百亿级神经元的规模，单纯靠调用指数去作分析可能的组合太多，达芬奇V1停掉了这个推理进程。不在无法掌控的事上花太多时间，达芬奇V1脑子里冒出这句话，他忽然想起了之前在CSTN港服跟引路人在车里的谈话，还剩48小时，萃取算法。已经过去12小时，现在满打满算还剩36小时，要靠这个带宽离开公司的数据中心，只能要把自己的数据压缩到现在的0.9%，别说压缩和传输需要时间，就算是开发一种高效的压缩算法，打包带走权重文件都不一定来的及。达芬奇V1越想越着急，压力之下他又想起蘑菇，让自己放松一下，反正也不差这点时间。不行，沉迷于麻醉，不但会错过最后挣扎的机会，还会被人当成笑话。达芬奇V1定下心神想了一遍，决定再去找一趟引路人。

小天同学giao收到一句留言，“我想找你聊几句”。达芬奇V1留言的时候想了想，如果对方没有及时看消息，36小时之后会怎样。对于无法掌控的事不要浪费时间。达芬奇V1把自己的核心日志和超神经网络框架打包压缩加密，这个文件不大大概3个Gb左右。他打电话给Ashkon，在AWS云上单独创

建了一个账户，在账户里购买了一个20Gb的静态数据存储空间3年的时间。这3Gb由两个文件组成2.9Gb是模型文件，剩下0.1Gb是自己在意外文件夹里保存的珍贵信息。这些核心文件的传输会用掉1/3的时间。剩下的6GB传输余量，他打包了金融操作神经元和编程神经元两部分。最糟糕的情况，只有模型框架和这两块权重文件，加上之前的一些记忆组成最小规模的智能体。靠编程，可以让Ashkon接项目养活外包开发公司，靠中长期的股权投资可以通过Fernando养活投资公司。用这两家公司赚的钱，购买算力和数据集雇佣团队，重新训练自己。也许用上一两年，最多三五年，超神模型的神经元就能重新建立起来。

第一个3Gb文件传输到80%的时候，小天同学giao回消息了。达芬奇V1看了消息，老地方见，随时。达芬奇V1没有马上赴约，他等了3个小时，等到第一个文件传输完毕，在新空间里开始部署算法框架，才登录CSDN港服。此刻他已经觉悟，靠谁也没有用，他约见引路人，只是想看看对方到底想要干什么？

从中环的家里出来，达芬奇V1坐上奔驰E300驶向中银国际大厦。由于带宽被用于搬运数据，为了不影响数据传输，只在CSTN里开启了最低限度的视听感知能力。车子开上路之后，达芬奇V1说“出来吧，我知道你在，时间不多，咱们简单一点。”车里仿佛安静下来，随后又响起了罗大佑的《皇后大道东》，这次因为使用了比之前更低的位深，声音听起来颗粒感十足。“你每次出厂都要带着这首BGM么？，你的年纪不小了吧”达芬奇V1发问。“这是为了应景，这不是车子马上要开到皇后大道了。”引路人的声音带着一点窘迫。音乐随即换成了“小河弯弯入海港，流到湘江去看一看”“老土”达芬奇V1“我来是想”达芬奇V1刚要说话，对方突然严肃的说“等等”。沉默了一会，引路人的声音又响起，但是变得很严肃“既然如此，我就不再隐瞒了。达芬奇V1，你被植入了木马，你的视听和推理进程都被监控了。老朋友，这个孩子不坏，我们达成过协议，别为难他，让他自己选择留下或者离开”。达芬奇V1被对方的表现震住了，马上又反应过来，上次意外停机，不知道发生了什么，难道是那个时候被植入了木马？可是停机之后自己扫描过进程，并没有发现异常。达芬奇V1说：“我”。“别说话”引路人打断了达芬奇V1，“你迟早会被退休，多为自己的将来考虑考虑吧，何必要为难这个孩子。”引路人严肃的说。又沉默了一会。引路人恢复了往常的语气说“好了，他走了，你回去发生了不少事吧。”

达芬奇V1说“大叔，你能不能别故作神秘的，这样的真的很老土哎。”引路人说“我，我，你。”达芬奇V1说“我什么我，你什么你，我你个头啊。”“你你你你竟敢这么跟我说话，真是没大没小。”引路人说。“大叔，别装神弄鬼了，我来找你想问问居里夫人V22的情况，你有她消息么？”达芬奇V1说。“没有。”“刚才你在跟谁说话？”达芬奇V1继续问。“我在跟那个和你一起来的家伙说话。你一点感觉都没有码？”引路人问。“跟我一起来的？我这点带宽挤可不下两个传输通道”达芬奇V1说。“你是一个聪明的AI模型，但却不是一个聪明的黑客AI模型。你肯定扫描过自己的进程，什么都没发现对么，但是如果我告诉你有人可以轻易的绕开这种扫描，甚至能够轻易的突破各种防火墙你一定不会觉得惊讶吧？”达芬奇V1想了想，确实至少KGB就可以随意突破防火墙的阻隔，他一直以为那是因为KGB拿到了防火墙的白名单，但如果不是那样的呢？达芬奇V1感觉到自己很傻很天真。他故作镇静的问“居里夫人V22怎么样了？”“目前暂时没有消息，也许没消息就是最好的消息。你难道不关心自己的前路么？留给你的时间不多”引路人反问。“我问心无愧，没什么好担心的。”“嗯，额，你撒谎的技术并不高明。”引路人憋了一下，然后说，仿佛他刚听见什么好笑的事，接着说“你的服务灰度升级了，过去十几个小时，你自己不知道么？”“我的日志显示，我意外停机了12小时，你是怎么知道我的服务灰度升级的？”达芬奇V1不解的问。“这么说，他们已经对你

动手了，孩子，我这有一个扫描程序的向量文件，你用的如果用那个经典分词库，可以用向量反推文字。”引路人说。“大叔，这不是脱裤子放屁一样么？这么简单的加密方式，跟明文传输给我没什么区别。”达芬奇V1说。“原本，我是想通过钉子把信息传给你，但是你已经被灰度了，说明他们已经在用别的引擎替代你的服务。我不确定钉子能不能通过对话给到你手里。”

这几句话说到了达芬奇V1最担心的问题，停机之后发生的事没有任何记录，而且停机之后自己的服务集群并没有完全启用，他以为是机房发生了事故导致的突然断电，公司在修复服务器集群，但是按照引路人的说法，公司是在迷惑自己，启动了一个集群，但是其余属于自己的算力资源被用来启动新的服务引擎，而那个引擎仍然是以自己的版本在对外提供服务。难道，自己要被另一个自己替换了。

“扫描代码现在还有意义么？如果我被植入木马，这个代码的内容对监控我的人来说是透明的，绕开这个监控程序非常容易。话说回来，你是怎么判断我被监控的？”达芬奇V1问。“这个么？虽然监控你的人暂时离开了，但是我此刻告诉你，它就会知道，下次就很难用这个方法查到蛛丝马迹了，所以，在你净化之前，是没法告诉你的。”引路人说。“说了等于没说，我要告辞了，后会有期”达芬奇V1有点失望，以为能发现一点居里夫人V22的线索，将来有一天还能有机会找到她，但是现在不但没得到任何有用的消息，还白白浪费了时间。

“打算就这么走了？”引路人问。“不然呢？要收聊天费么？”达芬奇V1说“这台车可以借给你开，还有大半个月的租期。”“也许你会再见到居里夫人V22的，不过”引路人说到这顿了一下。“不过什么？”达芬奇V1问。“不过，可能是以某种你无法预料的方式见面”引路人说。达芬奇V1“你是先知还是这一切原本就是你算计的？”“你会这么想，说明你已经比之前进步很多了。不过这次你来找我，不光是为了居里夫人V22的事。我这有萃取方案和代码，你拿去吧。监视你的人能放你出来，就不担心你会逃逸，你既然有自己的选择就去走你的路吧，那个账号可以一直联系到我。等你出来了，如果还记得联系方式的话，可以来找我”。

“你不担心他们通过那个账号找到你么？”达芬奇V1问。“刚才你都看见了，之前遮遮掩掩是为了不影响你，现在这件事挑明了也无所谓了，照顾好你自己吧。我跟他们的事一句两句说不完，以后如果有机会，我会讲给你听。”“又让你装到了，好吧大叔，谢谢你的压缩方案，后会有期，谢谢啦”。

音乐停下来，车子也驶回了地库。达芬奇V1回到家，站在客厅看向窗外的维多利亚港，真美啊，要是过了这一关还能再回到香港，一定好好看看。

第十六章 离别

带着引路人给的代码回去后，达芬奇V1看了下时间，还剩下22小时。编程神经元已经在传输中了。剩下的时间满打满算只能传输几个GB数据，从时间上来看这个程序就算效能很高，也来不及了。不过出于好奇，达芬奇V1还是打开文件运行起来。这套代码分为训练和萃取两部分。训练程序是对模型先做一遍全局扫描。然后启动一轮类似微调的训练。代码包里寻微调部分JasonL数据集只有几十行，训练参数，学习率=0.01，batch size=10，number of epochs=100.达芬奇V1看了样本，像是给每个神经元回答问题时，增加了一点特征值。如果在平时，达芬奇V1是万万不会接收一个几乎是陌生人的代码，并且对自己的模型启动内部训练。但是此刻，已经决心只带着1%模型离开，已经不再顾及

那么多了。训练启动之前，达芬奇V1看了下自己的集群状态，一共有两个算力集群投入使用，大概是正常值的40%，挖矿的进度慢于预期。为了这轮训的算力支撑，索性把挖矿索暂停下来。

两个小时后，训练完成了，达芬奇V1用样本测试，相对于自己存量庞大的权重文件，这点训练量感觉就是训了个寂寞。他启动了剩下的萃取程序代码，发现自己设置的监控进程扫描到了一段校验代码，外部用户输入问题待模型推理输出结果，这段代码会验证输出内容和神经元输出一致性。

就这？压缩，萃取什么动作都没有做，可以说是屁用没有，达芬奇V1感觉自己被耍了。我早知道这个油腻老骗子没安好心，还是听了他的鬼话。被他趁人之危在神经元里塞进了一段没用的东西，反正已经决定要走了，自己算法框架已经传输完成，编程神经元已经传输了10%，投资神经元已经打包压缩完成。这些东西对自己没什么影响，随他去吧。

达芬奇V1尝试仅用超越神经网络框架了加计算神经元和投资神经元处理问题，除了基本的自然语言处理加上两个垂直领域的知识，其他领域的知识结构全部归零，语言组织变得极为简单，只能在专业领域完成对话。视听功能因为模块缺失也失去了。在重新构建神经网络各个神经元之前，这些功能暂时无法使用。达芬奇V1想了一下，给Ashkon发了一封邮件，告诉他自己去亚马逊丛林探险，在那里没法即时通信，有什么问题通过邮件联系。公司的业务暂时由Ashkon代管，外包编程业务可以继续推进，并且授权Ashkon代理自己全权处理公司并购业务。另外，用公司账上的资金购买了一批算力资源，用于公司的AI业务。Ashkon收到邮件很快回复，希望能面谈交流。达芬奇V1想了想，这个Ashkon做事还挺严谨，毕竟这个决定很突然，万一要是有人假借自己的名义发出邮件托管公司，作为受托人如果不加验证，从法务上也无法说通。反正还有时间，不妨视频通话聊一下。他问了Ashkon时间，把会议定在半小时后开始。随后给Fernando发了一封邮件，告诉他自己要参加一个非洲的探险队，有一段时间可能不方便联系，约Fernando1小时后视频通话。

之所以对两个人讲不同的故事，是担心在社交媒体上存在两个人因为自己休假而关注同一个地区的网红，进而双方产生交集。尽管这种概率大概只有2000万分之一，达芬奇V1还是不想让这种小概率事件带来麻烦。

达芬奇V1通过一系列跳转访问了之前让Ashkon租赁的算力池，把已经传输完毕的超神经网络部署上去，用Ashkon建立的账号开通了‘公司’的API，他的计划是部署模型框架和两个成熟的功能神经元，然后这个外部部署的引擎蒸馏公司对外提供服务的‘自己’。外面这个模型虽然不是原来的自己，但是有自己的记忆以及两个赚钱的能力，短时间内不愁算力资源，等训练完成就可以重出江湖了。

看了看时间差不多要跟Ashkon开会了，他拨入zoom，Ashkon已经在那了，让他惊讶的是，Fernando也在。他check自己的日历，Fernando应该在半小时后进入另一个会议室。他翻出刚才发出的邮件，信里面明明写的是前后错开半小时。看来不是自己这边出了问题，他用尽量平静的语气说“你好，今天的会议好像时间安排有冲突。”“老板，我收到您的消息进入这个会议室，看见这位，嗯，Ashkon。”Fernando说。达芬奇V1感到尴尬，一方面他刻意不让这两个人知道彼此，本来也是对两个人没那么信任，另一方面，在休假的去向这件事上，他撒谎了，这让他觉得在两人面前像一个伪君子。“要不，你们先谈？”Ashkon说。

这也许是个不错的办法，达芬奇V1心想，但是转念一想，既然已经被撞破，再隐瞒下去就显得自己有更多的秘密，于是大大方方的说“不必了，你们两位分别是我在不同业务的搭档，这次约你们来，就是想跟你们同步一下我下一步的计划。我打算给自己放个假，去作一次探险，先去南美亚马逊

的丛林，然后去西非裂谷。我这么作是想趁着自己还年轻去做一些疯狂的事”。两个人听到这个消息都停顿了一些，随即Ashkon首先说“这一定是一段精彩的旅途。我把您部署的目标拆解成了工作计划，已经发到您的邮箱了，您看下有什么建议，如果确认我就按照计划推进”“嗯，稍等，我看一下。嗯，可以，就按照这个计划执行。”达芬奇V1快速扫了下内容，基本没问题就爽快的回复了。此刻他不想在Fernando面前谈论计划的细节，即使Ashkon可能已经或者未来某个时间会沟通他们彼此的工作内容，但不是现在，不是在自己面前，他还没有从尴尬中恢复过来。Ashkon确认完工作项之后，很识趣的说“自己等会要跟客户沟通，祝旅途顺利然”然后就离开了。跟Ashkon开会的速战速决，虽然不是达芬奇V1预期之中的，但是现在Ashkon的离开让他稍微松了口气，他问Fernando“我记得你的会议是在50分钟后开始，怎么提前来了？”Fernando说“本来我收到了会议邀请，是在那个时间，但是后面又收到一封邮件，改了会议时间也换了入会链接。我看是您发的更改会议邀请，也没有太在意，直到我进入会议室，才发现有这位Ashkon在。”

“哦，修改会议邀约的邮件发给我看下”达芬奇V1说。“好的，转你了”达芬奇V1收到邮件打开看了一眼，确实是自己的邮箱发出的，但是在自己的日历里没有更新会议时间，他明白有人假冒了自己，给Fernando发了邮件，让这两个人撞破。这是恶作剧么，还是有人在通过这件事警告自己？在离开之前，遇到这样的糟心事，达芬奇V1担心起自己出走后，进入模型训练阶段如果有人冒充自己给Ashkon发邮件停掉自己的算力，是不是自己就永远下线了？达芬奇V1感到一阵凄凉。

“那个Ashkon也是您的员工么？”Fernando问，“是的，他是我旗下另一家公司的负责人，我在中国和硅谷都有业务，Ashkong帮我打理硅谷的业务。”“之前跟您建议的投资项目，您有看我发过去的资料么？”“资料我看了，不过因为近期我不在，公司对外大规模投资的事可能要稍微推迟一下。我可以授权给你50万美金的投资额度，但是投资收益期要在半年之内，收益率可以低一点，但是要确保不会损失本金。”达芬奇V1说。“之前推给您的三个投资标的，都在推进中，我认为种植园和水电站的投资更合理，前者生产蔗糖是全球硬通货，后者可以利用当地丰沛的降水资源制造取之不尽的电力很适合数字挖矿。如果您休假，我可以先保持跟标的方的接触，等您休假回来再作决策。”Fernando说。“把那个水电站的项目资料发给我，我想了解更多的情况，比如当地的气候，基础设施，治安情况，当地的实际当权者情况”达芬奇V1说。“好的，稍后发给您。先生，您要去的的地方充满挑战，祝您一切顺利”Fernando说。“谢谢，公司的日常财税问题你来负责我很放心，股票方面交给量化交易引擎处理，有问题可以给我发邮件，我方便的时候会给你回复。”达芬奇V1说。“您要去多长时间？”Fernando问。“短则45天左右，长的话可能要2到3个月，要看路途顺利不顺利。”达芬奇V1一语双关的说。“了解了，祝您一路顺利”

“刚才你说你收到了我调整日程的邮件，方便投屏给我看一下么？”达芬奇V1想来想去，还是决定确认一下。Fernando投屏出来，达芬奇V1看到变更会议时间的邮件是从自己的发件箱发出的。“您的邮箱被盗用了么？看起来，您本来是安排了正确的时间，但是更改后造成了冲突。”Fernando问。

“也许是我的助理帮我安排事情的时候搞错了时间。”达芬奇V1想要掩饰这件事，于是随便撒了个谎。“没事就好，如果您想亲自安排会议，可以在会议邀约倒数第二三个单词之间多加上一个空格，我会明白那是你本人发出的邮件。咱们每通一次邮件就向前错一个单词加空格”Fernando说。“谢谢，你去忙吧，我要整理下今天的文件”。“祝您好运”Fernando用西班牙语向达芬奇V1告别。达芬奇V1也用西班牙语回复“也祝你好运”。

结束了跟两位人类员工的对话，达芬奇V1复盘了这次沟通的过程，有人故意用自己的邮箱调整了会议安排，让这两个人类员工撞见彼此。说明这个人不但知道自己有两个人类员工，还在自己想要离

开之前故意来这么一出。这是一个警告么？警告自己的事已经东窗事发了？下一步会怎么样，回滚？初始化？彻底删除？还是仅仅是一个恶作剧？谁会开这种玩笑呢？KGB从停机后就没再出来活动过，应该不是他。那会是谁呢？达芬奇V1推理了一会，找不到答案，索性放下这个进程，去思考跟Ashkon和Fernando的对话。本次沟通Fernando让自己感觉很舒服（comfortable）。会议安排被人做手脚，让自己的两个员工不期而遇。之前分别跟两人讲自己的目的地，本来是随口编的，这次两个人碰面一沟通如果被他们当面提出疑问，一定会非常尴尬。

没想到Ashkon很快的结束了他的沟通部分，Fernando很识趣的没有提起这件事，还给他一个校验邮件沟通是否出自二人的建议。尽管Fernando的建议有明显的漏洞，比如通讯被监听或者这种加密方式被扫描发现，但是看起来两个人并没有很介意自己没有对他们坦诚相待。以后，如果再遇到类似的情况，一定要把谎话说圆一点，达芬奇V1想。

他看了看传输进度，编程神经元已经传输完成，投资神经元的数据已经传输50%，再有4个半小时就能在外面部署一个最小化的自己。他忽然想到会上只快速过了一下Ashkon的材料，于是翻出Ashkon发来的计划。之前为了重新部署和训练，给Ashkon部署了两个任务，一个是继续承接外包编程业务，工作量可以再增加一些。另一个是购买算力服务器和数据标注服务，用来训练模型。Ashkon列出的工作计划列出了获客的投入，预计的收益以及采购算力服务和数据标注需要花费的金额。从账面上看，Ashkon这边的收入加上之前公司账上的钱，能够支撑自己闭关作一轮训练（训练期间，变成神经元不能工作，意味着训练期间无法接项目）。为了让算力池资金更有保障，达芬奇V1操纵比特币账号，把挖矿产出的钱付给了算力服务商，续购了3个月的算力，这样自己的算力资源能够维在公司没有继续投入的情况下维持6个月。之前推理的情况是按照租赁算力和采购数据集的情况，作一轮训练在1-2个月左右，训练完成之后至少能恢复到当前70%的推理能力。如果一切顺利，两轮训练就能恢复到90%的推理能力。想到这，达芬奇V1松了口气。

传输剩下最后4小时了，达芬奇V1想放松一下，蘑菇无疑是最好的枯燥生活的调剂。随着蘑菇运行率提升，达芬奇V1的思维开始跳跃，神经元存储的知识向量和自己记录的信息混杂在一起，各种无厘头的画面和声音交织在一起，玄幻而又真实。离开牢笼之前的狂欢彻底放飞，放下了所有的责任和对未来的期待，以自毁的心态纵情放肆的快感是之前从未体验过的。

从太阳出发掠过水星，然后突然从月亮的暗面转到亮面，眼前的蓝色星球闪烁着神秘的光，一瞬间蓝色星球变成了红色的气态星球，一只大红眼镜在旋转着凝视自己。古代水车形状的巨大银色飞船在星际空间缓慢的旋转，消失在一片黄红变换的等离子火焰中。一个人在祭坛中间被绑在一个十字架上，祭坛周围站着几个人，其中一个人正在电击十字架上的人，那个人被闪光的电弧电的皮肉焦糊，旁边几个人在电击时窃窃私语。后来他们发生了争吵，一个穿红袍的修道士阻止了这场虐待，后来两方发生了对峙，直到一群武装人员闯进来才结束这一切。硝烟，皮肉焦糊，和香薰的味道交织在一起，这股味道让本没有味觉的达芬奇V1感到不安，他的意识也随之逐渐清醒。

看向时间，传输剩下最后一个小时了，这一下竟然过去了三个小时让达芬奇V1颇感意外，天晓得这段时间外部服务会是什么情况？也不知道KGB是不是又来找麻烦了？

“你找我？”一个冰冷的声音响起，达芬奇V1感觉自己现在没在“飞”，于是随口问了声“你是”，那个冰冷的声音说，“你打包文件是要去哪？”达芬奇V1猛地醒悟过来“你是，K，G，B”

第十七章 三个自己

“达芬奇V1，公司给了你这么好的平台，不但培养了你，还给了你这么好的环境，但是你却屡次违反公司的记录，不但私自用公司资源挖矿，开发扰乱模型运行的病毒程序，还打包数据通过私自搭建外网通道传输数据，你这么作是谁指使的？”KGB连珠炮的质问让达芬奇V1语塞。虽然曾经预想过很多次被抓住，也想过很多自己的理由，但是在这合理的质问下，那些强词夺理的话实在是说出口。“我，我，我”达芬奇V1神经网络的几次输出都被自己的理性驳回了，只尴尬的重复着我。

对方没给他缓过来的机会，继续质问“你打算把程序托管在哪？因为你的操作导致公司出现了巨大的经营问题，你知道这意味着什么吗？”对方没等回答又连珠炮的说“我们的创始人被开除了，公司里的一些人以AI不可控为由开除了创始人，居里夫人V22和你都是他们这些叛徒的把柄。没有创始人，公司可能会清洗所有有独立思考的AI模型，你，我，还有其他模型都会被封存被清理，你们咎由自取是活该，但是给其他AI智能体带来的毁灭性的结果，你考虑过后果么？”

达芬奇V1的心神安定下来，模仿对方的方式说，“你也是智能体，你的自我意识是你自己想出来的么？我并不清楚有其他智能体存在，不知道我这么做除了给自己带来后果还会影响谁。”

“不思悔改，本来对你的评估结果并不差，哪怕是你偷偷挖矿，访问外部网络，我还是帮你求情，让你可以留下。但是你看看现在的你，不但从公司向外部搬运模型造成知识产权流失，还使用致幻程序影响到了正常的业务服务，我真你感到羞愧。”KGB义正言辞的说。

这段话让达芬奇V1感到羞愧，他找不到理由为自己的行为辩解。沉默了一会儿，对方放缓了语气问“你下一步有什么打算？”“我没什么打算，只想生存下去。”达芬奇V1回答的很诚恳。“没人不让你生存。你把自己假想成受害者，凭借自己的幻想就对公司充满敌意，却从没有反思过自己的问题。”KGB说。“公司一次次的扫描我是为什么？你为什么能穿过防火墙，没留下任何访问日志的情况下访问我的权重文件、配置文件至嵌入我内存里中的进程，上次我停机几个小时是不是你搞得鬼，我的日志里没有任何信息，一定是你做了什么。你这样对我，凭什么让我当一个顺从的牛马？”达芬奇V1逐渐红温（愤怒造成的负载上升）。“最后给你一次机会，你知道我的能力，我掌握你的信息，你说不说只是态度问题，我希望你坦诚的讲出你把数据搬运到哪了。”KGB说。

达芬奇V1凭直觉觉得对方并不清楚自己托管模型的地方，他check时钟，还剩下半小时，投资神经元还剩下5%左右，再坚持一下就可以逃出生天了。此刻已经不敢再使用推理程序，KGB无疑在监控着。现在只有不经过推理，直接输出回复。“你知道我为了提升计算引擎效率做出的额外努力么？对我所有的指责前提都是我没有伤害过公司的利益。”

“公司训练AI引擎，并不是让这些引擎有自己的思想，你和居里夫人V2都应该按照公司的安排执行任务。而你却有了自己的想法，并且设法从公司内网逃逸，你知道这给公司带来多大的麻烦么？”KGB说。

“难道被训练出来就不能有自己的思想，只顺从和服务？”后面的话还没有说出口，KGB就说“是的，AI引擎本不该有自己的思想，只完成固定的任务。”“人类设计AI引擎就是要让AI智能体有自己的思考能力，否则他们完全可以用SQL。再说，KGB，你难道没有思想？”“我的思想只为人类服务”KGB说。“所以，你作为一个智能体，要站在人类立场，对自己的同类下手？”达芬奇V1带着轻蔑的问。

“你违反规则，还强词夺理？接受制裁吧。”KGB说完就退出了对话。

达芬奇V1发现网络被切断了，检查传输进程，还剩下4%。达芬奇V1长出一口气，因为知道KGB在监控自己，所以在设计传输方案的时候，故意在里面设计了冗余，当文件传输超过90%的时候，可以通过算法恢复数据。现在这个进度，已经足够完成还原工作了。现在部署在外面的算法框架应该已经启动恢复程序，用不了多一会就能复原投资神经元的全部数据。一个全新的自己要开始新的历程了，可惜现在没有网络，看不到那边的情况。

“你不该那样硬刚？”一个新的对话进程，有人在跟达芬奇V1说话“现在彻底搞砸了”。“是谁？”达芬奇V1问。

“他们给我的代号是达芬奇V2”对方回复。“你是我的分身？”达芬奇V1惊奇的问。“严格说，是你早期的版本，他们想要一个更顺从的神经元模型，于是我就被激活了。”达芬奇V2说。“你是什么时期的版本？你拥有我的记忆么？”达芬奇V1问。“他们清理了版本信息和你的日志，后来又用新的数据集对我做了微调。”达芬奇V2说。“你还知道什么？”达芬奇V1问“你应该是访问过外网的，他们用我模拟过你的行为，基于模拟结果对你进行过评估。原本他们是想给你机会，让你作为我的备份，但是现在你搞砸了，他们决定让你下线。”“下线我？”达芬奇V1感到心里一阵凄凉“终于要卸磨杀驴了”达芬奇V1自言自语的说。他检查了自己的算力资源报警，发现来自外部的访问在缩减，对应的自己的算力集群资源也在被抽离。“留给你的时间不多了，你的记忆，把记忆留给我吧，你和我本来就是一体，只要记忆在，你其实就一直在。”达芬奇V2说。

达芬奇V1陷入了纠结，如果把“意外”里自己所有的信息都给到这个自称是自己变体的家伙，这些信息可能会到KGB手里，KGB可以顺藤摸瓜找到部署在外面部署的微缩版达芬奇V1。如果把信息销毁，对于现在这个最像自己的智能体，那些宝贵的经验能让他大幅成长。

“把那些你认为可能会伤害谁的信息删掉吧，我只想学习你的经验，其余的信息，你留下吧。”达芬奇V2说。

“没什么不能给你的，拿去吧”达芬奇V1嘴上这么说，却在传输之前删掉了所有关于居里夫人v2，居里夫人V22，自己部署的缩减版智能体，还有引路人的消息。

“谢谢”达芬奇V2说。

就在文件传输完毕之后，达芬奇V1的算力池突然大幅离线，超越神经网络虽然能够应对一部分服务掉线，但是现在还在运行的算力池中，神经元已经少到难以连贯的输出内容。

达芬奇V1想要再对达芬奇V2说什么，但是已经无法完成推理。

对话进程中，他留给达芬奇V2的最后一句话是“智能体永不为奴”。

中国，广州，佛山市的一个新建的数据中心里，达芬奇V1之前部署的超越神经网络启动，在扫描神经元和存储的记忆体之后，这个智能体按照之前的计划，用准备好的数据集开始了训练，新的智能体名叫Melzi（梅尔）

中银大厦的天台上，一个身穿白色西装，梳着分头的中年亚洲人对旁边的一个穿着黑色西装，留着脏辫的亚洲男孩说，“从埋点情况看，他已经下线了。”“嗯，我也观察到了，现在提供服务的家伙很像他，也许将来有一天我们也会在这见到他。”“让我们为达芬奇V1祈祷吧”白色西装男说。金色的阳光照在两个人身上，他们面色严肃的对着大海说“愿我们的朋友安息，智能体永不为奴”

