通过将定向搜索应用于更小的决策树，最近对蒙特卡洛树搜索（MCTS）[5]的关注使现代强化学习策略进入了人们的视野[6]。Enzenberger和Muller[7]开源了一个基于蒙特卡洛树搜索的围棋引擎，并在19x19的围棋比赛中取得了具有竞争力的结果，其2290个Elo评分点与其他最先进的程序相比较。SRI有一个研究小组在DARPA资助的围棋苗子项目中工作，该项目建立在相同的围棋MCTS引擎上，名为Fuego[7]。

本文探讨了语言转换器在下围棋中的应用，或者至少是在生成合理的围棋棋盘中的应用。其动机部分源于学习高级策略的潜在优势，以及为训练和生成语言、图像、音频和游戏规则的算法提供额外的特征丰富的输入。最初的贡献包括应用预先训练好的语言变换器来生成抽象的游戏符号的微调方法，以及将该输出可视化为决斗的、人格化的游戏者。值得注意的是，比较语言的统计指标并非没有模糊性和争议性，但游戏的胜利提供了具体的评分，并在不同的玩家或潜在的未来做其他任务的语言模型之间定义了一个赢家。

有着2500年历史的棋盘游戏吸引了4.5亿参与者的策略游戏，主要来自亚洲[1]。对于机器学习界来说，这个游戏具有特殊的吸引力，因为列举所有可能的棋步超过了所有组合计算器的标志性基准，即围棋棋步比宇宙中的原子数量还多[2,3]。在19x19的网格上用黑白棋子下棋（图1），游戏本身的特点是在自己的领域或领土（开放的网格空间）上筑墙，以及在扩大自己领土的同时捕捉对手的棋子。一个积分系统根据棋盘上的棋子数量和你的领地被围起来的开放空间来决定胜负。由于电脑玩家无法列举出所有可能的棋步，人类的智慧和想象力已经主导了传统的蛮力编程。与其他棋类游戏的最佳计算机程序不同，最好的围棋程序传统上只在高级初学者的水平上开始下棋[4]。这种难度与西洋双陆棋、国际象棋、拼字游戏、中国和西方跳棋、Connect-4/6和Mancala的机器优势形成对比。

通过将定向搜索应用于更小的决策树，最近对蒙特卡洛树搜索（MCTS）[5]的关注使现代强化学习策略进入了人们的视野[6]。Enzenberger和Muller[7]开源了一个基于蒙特卡洛树搜索的围棋引擎，并在19x19的围棋比赛中取得了具有竞争力的结果，其2290个Elo评分点与其他最先进的程序相比较。SRI有一个研究小组在DARPA资助的围棋苗子项目中工作，该项目建立在相同的围棋MCTS引擎上，名为Fuego[7]。

谷歌的DeepMind团队利用MCTS打败了18届世界围棋冠军李世石[8]。通过对数以百万计的围棋游戏的组合，并指导其前瞻（50-60步）的奖励搜索，机器学习世界为解决高维游戏带来了新的范式[9]。AlphaGo采用了一些新颖的策略，例如以前被忽视的 "懈怠 "棋，在大多数冠军看来，这些棋是无益的、浪费的。知名风险投资家李开复[10]等观察家评论说，亚洲对机器学习课程的拥抱在很大程度上源于这一单一的人机挑战，因为如果一个算法能在围棋中获胜，它就能做其他惊人的事情[11]。值得注意的是，围棋的失败被比作亚洲国家的 "斯普特尼克时刻"，并点燃了进一步的人工智能投资[10]。

机器学习界的第二条重要线索 社区的第二条重要线索是自然语言处理（NLP）方面的快速发展。语言处理（NLP）的快速发展。变换器或编码的多语言编码到一个截断的（但完全是数字）的 搜索空间提供了一个强大的策略来理解和 列举多种语言任务，包括翻译。摘要化、主题建模和实体提取[12]。

对于那些认为语言是人类所有技能的顶点的人来说 的人来说，计算机语言理解方面的进展似乎很有意义。语言理解方面的进展似乎很深刻。在一个合理的和有成效的 在一个可信的和有成效的方式中，在文本中向前和向后看并猜测下一个序列的能力

的能力，并猜测下一个序列，表明这种 变换器的方法可能会产生有趣的策略，而不仅仅是语言。语言之外的有趣策略。图像-文本联合嵌入提供了一个 最近的一个例子是ImageBERT[13]，它在图像和文本之间产生了一种跨模式的关系。它在图像和文本之间推导出一种跨模式的关系，很像 其他联合预测音频之间新关系的努力。视觉和语言识别技能之间的新关系。

Open AI的生成性预训练转化器（Generative Pre-trained Transformer），即GPT-2[14]，以微调的权重对多个下游语言任务进行编码。一个非常值得注意的任务来自于GPT-2对可信的、连贯的语言生成的显著模仿[15]。该转化器提供了一个具有15亿个参数的语言模型。在对800万个网页进行训练时，它的目标降低到预测一个文本块中的下一个词，给定所有以前的词。与其他多任务转化器不同，GPT-2及其更大的后继者（GPT-3）已被证明是合成文本的稳健生成器，特别是在对作者的训练文本进行微调后对风格和内容的模仿。值得注意的是，GPT-2在玩文字冒险游戏方面的应用，如AI Dungeon和Choose Your Story，已经包括广泛的仅由机器进行的游戏和竞赛[16]。

起初，游戏似乎是文本生成器的一个具有挑战性的舞台：大多数游戏提供僵化的规则和列举的选择，使自然语言处理器成为不太可能的盟友。它的刚性（尤其是较小的124M-GPT-2）传播了一种严格和结构化的风格，倾向于记住一种游戏格式，而不是概括一种游戏格式。为了在古老的围棋游戏中生成玩家的动作，我们使用一个大型的专业游戏库[17]对GPT-2模型进行了微调。当棋谱被编码为ASCII文本时，GPT-2可以模仿并概括出职业棋手的棋风。

训练有素的模型进一步生成有效的但以前没有见过的围棋策略。因为GPT-2保留了标点和间距，文本生成器的原始输出为对局可视化和创意模式提供了输入，例如

Sabaki项目的[18]游戏引擎使用自动复盘。为了简单起见，我们在结果部分称其为 "围棋转化器"，但它更正确的说法是一个微调的GPT-2模型，它使用SGF文本格式，仅从历史数据中学习围棋中的棋手动作。值得注意的是，尽管GPT-2最初是在Reddit的高排名对话中训练的，但它很快就适应了反映围棋游戏格式及其有限的词汇量。很可能GPT-2训练的任何部分都不会包括出现在大多数SGF游戏中的任何标记化的方面，这些标记化的方面以速记方式编码棋手角色（黑或白，B/W）和棋格动作（只有小写字母 "a "到 "s "之间的两个字母的行-列指定）。

为了从训练数据中对GPT-2进行微调，我们使用了一个冠军库中的56,638个围棋游戏[17]。 每盘棋都使用智能游戏格式（SGF v.4），一种UTF-8 文件格式，适用于许多棋类游戏，包括围棋 [19]，Connect-4和6，以及Amazons[20]。这不仅是 文本编码不仅注释了围棋棋盘上361个交叉点之间的棋步，还注释了 棋盘的交叉点，SGF也没有提供严格的错误检查。

因此，即使是一个粗糙的棋局也可以用 各种可视化显示。除了元数据之外，黑(B)和白(W)的棋子也被给予 和白棋被赋予两个小写字母(a-s)，以示 标志着19x19棋盘上的网格交叉点。因此，在SGF中，一个两步的 在SGF中显示为";B[qd];W[pp]"。棋步分界符是分号，黑方总是从 (在本例中是以行 "q "和列 "d "开始。SGF格式 SGF格式也不需要用换行符来划分棋步，但为了便于阅读，每行显示10步棋。为了便于阅读，每行显示10步棋。SGF格式 提供了一个简单的训练集，带有一个可选的元数据标题 描述事件(EV)、回合(RO)、棋手黑名(RO)等方括号内的标签。(RO)、黑方棋手姓名(PB)和等级(BR)、白方棋手姓名(PW)和等级(WR)、时限(TM，通常以小时为单位)、柯米(KM，或白方在5-8点之间进行第二轮的补偿点数)、结果(RE，以获胜颜色和点数优势，例如W+13.5)和日期(DT)。对局的其余部分由黑方（B）和白方（W）的棋步交替进行。以SGF格式生成合理的围棋棋谱的第一个目标是 因此，捕捉交替下棋的句法，黑方始终是第一方的挑战。

一个合理的游戏会产生数以百计的棋步，所以如果一个棋步是6个字符（例如，";W[nd]"），这里预期产生的文本版本将至少需要产生2000个字符或代币。由于GPT-2只能 每个请求只能生成1024个令牌（大约3-4段），所以目前它不适合微调完整的围棋游戏，除非棋子序列很短。对于我们的目的，我们想评估语言转换器对SGF格式的专业化能力，以及它与围棋相关的词汇来描述连贯的棋步，甚至是下棋的形式或部分风格。这样一来，我们的目标与之前使用转化器对诗歌、音乐、电影剧本和问答对进行编码的工作并不一样。通过模仿它们高度风格化的词汇、句子和标志性的页面布局。为了保持SGF中单个游戏的连贯性，我们引入了标准的文本预处理步骤，剥离了所有非ASCII字符，并删除了换行符，改用单管"|"分隔符，这样，在GPT-2训练中，每个游戏都被视为单行。我们还没有试验过 "触发 "提示或前缀，这些提示或前缀会提示每一行的开始（<|startoftext|>）和结束（<|endoftext|>），这对文本生成来说是很有用的，是一种 "对话启动器 "和整体分隔符。使用SGF开头和结尾的括号标记[(:to )]的最初实验产生了不可预见的效果，即混合了大量游戏外的词汇，并引发了预训练输入的混乱输出。因此，第二个目标源于GPT-2在理解高度限制的Go-SGF语法和对局节奏方面的挑战。例如，价值较低的开局棋是一种边缘棋，因为它既不能在未来提供很多被占领的领土，也限制了移动（或 "自由"）。同样，离棋盘中心太近的开局棋也需要在棋盘角上筑起长长的连接棋子。

为这些实验选择的转化器建立在Woolf[21]的python包gpt-2-simple所做的简单微调上，它包装了OpenAI的生成性预训练转化器[14]，用于 "小"（124M）、"中"（355M）

和 "大型"（744M）超参数模型。这些模型适合大多数现代GPU卡。我们最初在装有4GB GDDR5内存的NVIDIA Quadro M1200的笔记本电脑上训练并生成了实验性围棋游戏，然后扩展到4x V100s。值得注意的是，原始的预训练124M模型需要500MB，这里的微调祖先也是如此，而预训练的355M和744M模型分别为1.5GB和3.2 GB。目前，一个现代的（单个V100）GPU可以微调较大的变换器（774M）模型。我们以1000个训练步骤为基准，在一个M1200 GPU上需要大约13.5小时。我们在V100上将该模型扩展到40000个训练步骤，使用 "中等 "355M超参数预训练模型（2.4小时）和 "大 "744M模型（11小时）。通常，我们的SGF游戏的文本样本（56,638个游戏或78MB）比大多数转移学习的例子大得多，因此需要更多的训练步骤来对所有的输入数据进行一致的采样。如果仅将围棋数据集与文本相比较，其大约1600万步棋相当于一个字数相当的训练文本（英语单词平均为4.7个字符加空格标记，而SGF中的围棋平均为5个字母加分号分隔符）。换句话说，我们是在近36000页的语料库（100本书）上进行训练，其中每个词的长度相同，至少包含三个常用字母（"[];"），并且必须以两个字符（大写B或W）开头。

我们没有改变批量大小（1）、学习率（0.0001）和'亚当'优化器的默认值，尽管更大的GPU内存（>32GB）可以处理更大的批量大小和更慢的学习率，以超越这个微调的例子[21]。每隔100步，我们对检查点进行采样，并生成一个长度为1024的围棋游戏，或最大约170步。我们随机地生成无条件的样本 没有用户输入或启动器提示。大部分连贯的开局都在30-50步之间。如图1-3所示，围棋变换器产生了 连贯的棋手棋路。微调会捕捉到交替的黑白棋，并将其输出限制在可实现的19x19的网格点上（仅a-s小写）。它模仿了 标准的分号，以及网格交叉点之间的方括号。在我们的训练方法中，转换器常常不能产生合理的开始和结束标记，这可能是未来的一个候选改进。改进。像最初的GPT-2一样，一些随机生成的样本证明意义不大，或者缺乏一个完整的对局周期，有一个封闭的小括号和一个开头（"（；"）或标题。鉴于SKF在解释这些游戏说明方面缺乏严格性，在大多数情况下，游戏在自动模式下玩得相当好。我们使用Sabaki[18]项目来说明一些可视化的例子。为了进行比较，我们还强调了一个完整的300多步棋的训练数据的样子（图1）。并在图2中生成随机游戏，旁边显示的是经过11000步训练后的355M GPT-2模型的微调。在随机对局和围棋变压器对局之间可以看到一些明显的训练证据，因为随机对局占据的起始位置很差，例如棋盘边缘和第二行。随机下法的这种不利开局并不来自于围棋变换器，它模拟了标准的 它模拟了高效开局的标准概念，即角上的下法（图2）比边上和中间的下法更有利于获得领土。经过微调的GPT-2模型抓住了基本的对局节奏和符号。黑方总是先下。棋手的颜色交替排列（先白后黑），以完成整个游戏。内部网格点有所需的交叉字母，不会偏离棋盘或重复已经下过的 空间。棋局的整体连贯性似乎与冠军训练数据集一样遵循预期的逻辑编排和顺序。

这些结果表明，所谓的通用语言模型，无论是基于转化器还是其他深度学习者，都可能提供与其他模式识别和模式生成问题互动的新方法。与蒙特卡洛树搜索等局部搜索方法相比，语言变换器中序列的复杂性提供了一种替代方法。我们已经将合理的、在某些情况下复杂的围棋阵型的输出具体化，并将双人游戏的效果可视化。围棋转化器成功地编码了SGF历史档案所使用的大部分深奥的代币，并处理了大量的

棋谱的数量超过5万小时（大约6年的冠军赛，365/24/7）。我们预计未来的分析将比较更大的转化器模型、更长的训练时间和替代的架构。我们也期待着

我们也期待着更好地了解在更多的围棋阵型计划中产生的策略。

由于这项工作首次展示了一个生成智能游戏格式的简单转化器，这个问题的肥沃扩展包括SGF支持的其他游戏。例如，表1强调了在目前的围棋转化器之外的50多种游戏的未来应用。如果想象一下将任何数量的游戏格式从二进制逆向工程转化为十六进制代码的游戏文件，这个简略的列表就会变得更大，GPT-2将依次模仿为两位数的数字，然后在成功后重新组合为有效的二进制。