LLMs 和代理 AI 在保险决策中的应用：非洲的机会与挑战

Graham Hill*, Jing-Yuan Gong*, Thulani Babeli‡ Moseli Mots'oehli†*, James Gachomo Wanjiku*,
*The Shard South Africa; †University of Hawai ʻi at Manoa; ‡Elenjical Solutions South Africa

摘要—在这项工作中，我们强调了人工智能（AI），特别是大型语言模型（LLMs）和代理型 AI 在保险领域的变革潜力。我们考虑并强调了在性能迅速提升、开源访问增加、部署成本降低以及 LLM 或代理型 AI 框架复杂性的情况下，保险领域面临的独特机遇、挑战及潜在路径。为了使其更贴近实际，我们识别了非洲保险市场中的关键差距，并突出了重要的本地努力、参与者和合作机会。最后，我们呼吁精算师、保险公司、监管机构和技术领导者共同努力，旨在为非洲人创造包容性、可持续性和公平的 AI 策略和解决方案：由非洲人创建并服务于非洲人。

***Index Terms*—非洲人工智能、保险、大型语言模型、代理型人工智能、精算科学、政策制定**

I. 介绍

人工智能（AI），特别是大型语言模型（LLMs）和编排代理 AI，正在改变金融行业，尤其是在保险公司分析风险、处理索赔以及做决策方面。然而，大多数的 AI 活动发生在过去，并且主要在非洲之外发生，这导致了非洲保险市场中未被解决的问题和机会仍然存在。为此，本文通过调研区域研究努力、初创企业和既有企业来探讨 LLMs 和代理 AI 如何为非洲保险公司创造价值。我们旨在激发一场鼓励精算师、保险专业人士、监管机构和技术领导者之间的合作的对话，以构建由 AI 驱动的保险解决方案，特别是那些仅对非洲市场可行且有益的解决方案。本文剩余部分将回顾关键的人工智能概念和全球用例，然后深入探讨非洲的举措与挑战。最后，我们将讨论一个很少被提及的技术采纳方面：新技术与监管之间的关系。这在高度受监管的金融市场中尤为重要，例如南非 [1]。

II. 大规模语言模型和代理人工智能在全球保险中的应用

在本节中，我们追踪了从语言建模（LM）和自然语言处理（NLP）到 LLMs 及其在代理 AI 角色中的演变。我们陈述了 LLMs 和代理 AI 在全球范围内在敏感且重要的 [3] 以及受到高度监管的保险行业 [4] 中的应用案例和挑战。

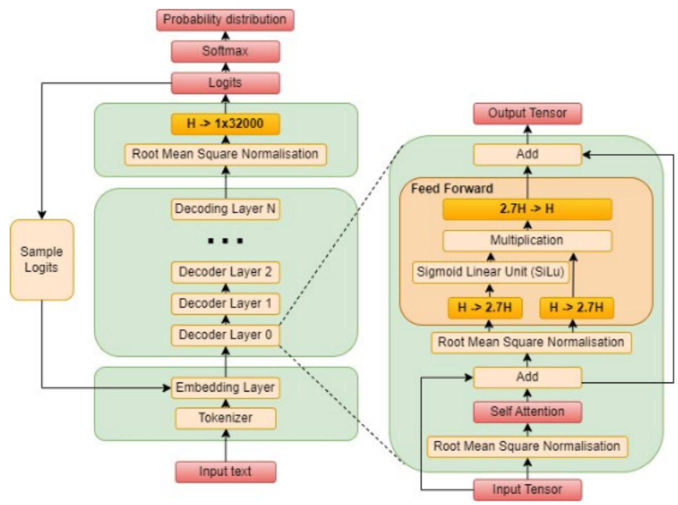

图 1: 一个示例变换器 LLM（LLaMa2）架构（改编自 [2]），通过堆叠的自注意力和前馈层端到端处理用户提示。

A. 大语言模型和代理框架

在大语言模型之前，我们有较小的语言模型 [5]。我们使它们更加复杂 [6], [7] 并提供了更多数据，但它们表现不佳且速度慢。随着计算机变得更快 [8]，我们使它们体积更大 [9]，给予它们更多数据 [10] 和计算能力，但几乎没有什么显著进展 [11]。然后，我们简化了它们 [12]，但增加了更多数据 [13] 和计算能力 [14]，最后它们运行正常 [15]。起初它们表现得很好，我们开始期望它们做得更多，但很快我们就意识到它们在理解一些困难的概念 [16] 甚至是某些基础概念 [17] 上遇到了问题。然而，通过精心设计的提示，大语言模型几乎可以瞬间

提供准确的答案，使它们成为需要快速响应的任务的理想选择 [18]。

迄今为止，最先进的（SOTA）模型是多模态变压器模型 [19]–[22]，通过诸如从人类反馈中进行强化学习 [23]、专家混合 [24]、思维链微调 [25]、知识蒸馏 [26] 以及最近的迭代宪法对齐 [27] 等方案训练而成。随后，这些模型被微调以专精于特定的知识领域：金融 [28]、保险 [29]、生物学 [30]、法律 [31]，以及语言：法语 [32]、西班牙语 [33]、中文 [34]，以及我们将要在第 III 节中探讨的多种非洲语言 [35]。与其让单一的大型语言模型超载，为各个模型分配专门的角色可以提升它们的实用性，将它们视为具有不同任务和能力 [36] 的代理，并通过一个编排器将其输出链式连接 [37] 以响应用户的提示。代理 AI 框架将用户提示分解成明确的规划阶段，以便生成并由专门的代理执行子目标，在系统内评估并改进这些响应，所有这些都是由一个中央控制器在返回给用户的可能是正确且合理的响应之前反复循环这些阶段来编排的。常见的代理 AI 框架和库包括 LangChain [38]、LlamaIndex [39]、Microsoft Semantic Kernel [40] 和 Auto-GPT [41]，所有这些都内置了对函数调用、内存管理和多步骤工作流编排的支持。

图 1 和 2 分别说明了 LLM 中将提示映射到响应的变压器块，以及一个代理系统，在该系统中，调度器将任务拆分为子任务，将它们分配给专门的代理，并将其输出合并为上下文感知响应。

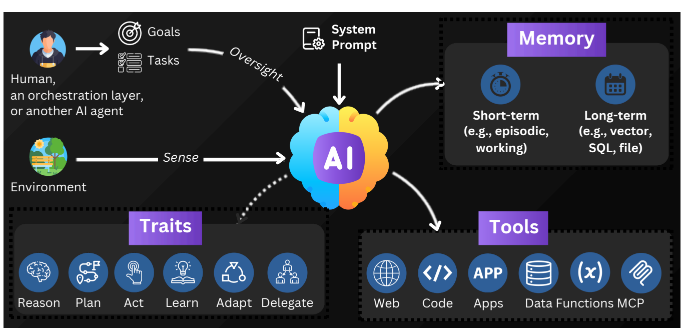

图 2: 代理 AI 系统（改编自 [42]）其中中央协调器将请求分解为需要不同工具和内存访问的子任务，将它们分配给具有不同特性的代理，并将其输出重新组合成一致且上下文感知的响应。

B. 全球用例与挑战

随着开源可用性、推理能力和实用性的提高，LLM 的 AI 代理索赔处理解决方案已在全球范围内实施。在日本，一家拥有 100 年历史的保险公司**福国相互生命保险公司**用 IBM Watson Explorer 的认知 AI 取代了多达 30 名理赔评估员。该 AI 具有自然语言理解和非结构化数据分析（文本、图像、音频）的能力，用于评估医疗证书并计算索赔赔付金额 [43]。虽然这一部署发生在现代 LLM 时代之前，但它带来了据报道的 30%生产率提升和每年节省 1.4 亿日元的效果，同时仍然依赖于每项索赔的人工审核批准，突显了早期 AI 在保险中的前景与局限性 [43]。**可处理.ai** 使用 AI 进行车辆和财产损害理赔评估。他们主要依靠卷积神经网络（CNN），而非 LLM；他们的解决方案可能通过视觉语言模型升级，以处理损坏图像及汽车品牌、型号和其他用户输入信息的文本描述，从而生成准确的维修估算。他们使用 Bi-LSTM [6] 模型进行光学字符识别（OCR），以从图像中自动提取关键文本 [44]。Tractable.ai 与 GEICO 合作，每年为包括 Tokio Marine 和 Mitsui Sumitomo（日本）、Covea（法国）、Aviva（英国）和 Admiral Seguros（西班牙）在内的 35 多家保险公司处理超过 20 亿美元的索赔，其中包括最大的汽车保险公司 [45], [46]。**辅助记忆的**：一家印度人工智能咨询公司，使用经过微调的大语言模型（LLM），如 OpenAI 的 GPT-4 进行政策分析、索赔问答，并为个性化客户支持聊天机器人提供动力 [47]。截至 2022 年，据报道 Aidetic 从其 20 多家客户那里获得了高达 100 万美元的收入。

技术转移，一家法国的保险公司 AI 平台，结合了微调后的 LLMs 和代理型 AI 以优化承保、理赔和欺诈检测过程中的决策 [48]。在 10 个国家设有办公室，为超过 100 家保险公司提供服务，并且仅在 2021 年就分析了高达 50 亿份的保险单、索赔案和文件。自 2020 年以来，该公司将 LLMs 与微调后的保险特定数据整合，以实现其理赔和欺诈工作流程中的文档分类、实体提取以及决策总结自动化。他们强调模型可解释性，并通过人工介入审计轨迹来提高稳健性，从而实现了 97%的端到端索赔自动化 [49]。虽然这是一个非详尽无遗的涵盖范围，我们还是在表 I 中列出了其他值得注意的应用案例并进行了分类。

尽管取得了这些进展，实施仍然面临挑战，因为保险公司需要在每个司法管辖区导航严格的个人数据隐私和治理要求。从技术角度来看，将现代 AI 系统与遗留 IT 基础设施 [50] 集成仍然是主要瓶颈之一，这一问题因非技术公司中具备足够技能的员工短缺而加剧，并

公司	国家	建立	任务/类别	人工智能技术
莱拉帕 AI	South Africa	2022	African language comprehension	Instruction-tuned LLMs
虹彩鸟	Kenya	2018	Health & life micro-insurance	LLM chatbots & AI validation
裸保险	South Africa	2017	Claims automation	GPT-3.5 - powered chatbot
菠萝	South Africa	2016	Claims processing	CNN vision + LLMs
普拉顾问公司	Kenya	2015	Parametric agri-insurance	Multimodal + LLMs
令人信服的	India	2020	Voice-based fraud detection	Speech-to-text & LLMs
记忆相关的	India	2018	Policy analysis & chatbots	Fine-tuned LLMs
索赔力	USA	2018	Claims & workflow automation	AI-assistant
被量子系统	UK	2017	Risk & capital modelling	Data-driven actuarial models
超指数分布	UK	2016	Pricing & underwriting optimization	ML-driven pricing
柠檬水	USA	2015	Claims processing	AI Jim - ML + NLP
赛托拉	UK	2015	Underwriting & risk scoring	LLMs & ML risk intelligence
技术转移	France	2014	Fraud & claims automation	Generative AI & LLMs
可处理的	UK	2014	Damage assessment	CNN vision + bi-LSTM OCR
FRISS（欺诈风险智能）	Netherlands	2012	Fraud detection	Generative AI
不安全性	USA	2001	Policy & claims management	Cloud-native AI modules
科学软件	USA	1989	Underwriting & risk assessment	LLMs
CCC 智能解决方案	USA	1980	Auto-estimate generation	Computer vision models
福国相互生命保险公司	Japan	1923	Claims assessment	IBM Watson Explorer cognitive AI

表 I: 选定的专注于保险的人工智能公司，它们的起源和成立年份、主要任务以及所基于的大语言模型或代理人工智能技术。Lelapa AI 在低资源环境下特别是在核心多模态非洲语言任务中尤为领先。

且还需要管理定制模型开发和维护的巨额成本 [51]。尽管存在潜在的好处，但当无法深入了解这些 AI 系统的内部工作原理时，获得利益相关者的信任和支持可能非常具有挑战性。没有真正的可解释性、透明度以及与 KPIs 的明确一致，采用速度已经非常缓慢。在下一节中，我们将探讨专门设计或实施用于非洲语言和文化背景下的 LLM 和代理型 AI 解决方案。

III. 非洲人工智能：研究、部署与保险

本节概述了非洲的人工智能生态系统，重点关注大型语言模型和自主人工智能。我们回顾了非洲各大高校、实验室和初创公司在大型语言模型和自主人工智能方面的学术与工业研究，然后探讨它们在保险领域的应用。

A. 非洲人工智能研究

在南非，早期的深度学习自然语言处理工作集中在单一语言中的个别任务上，例如：[52] 用于 isiZulu 形态分析和 [53] 用于 Setswana 掩码语言建模，生成可用于训练其他下游任务模型的嵌入。在尼日利亚，为约鲁巴语 [54]、伊博语 [55] 和豪萨语 [56] 构建了类似的掩码语言模型；这些模型随后被微调用于命名实体识别、情感分析和问答任务 [57]。在肯尼亚，研究人员 [58] 展示了双向 LSTM 对于斯瓦希里语命名实体识别的实用性。早期的工作表明我们可以超越 Google Translate 处理非洲语言的能力，但像 Masakhane [59] 这样的倡议通过协作研究带来了开放、整理的数据集、基准测试和各种规模的多语言模型。

至今为止，经过非洲语料库微调的多语言变压器模型如 AfriBERTa [35] 在多项任务中取得了具有竞争力的结果。在 [60] 中，作者对 GPT-3.5 和 LLaMA 变体进行了基准测试，在八种非洲语言上显示出低资源语言性能显著下降，这为定制的 Afrocentric 模型提供了论据。在 [61] 中，作者提出了一种框架，该框架在构建 LLM 时将非洲语言和文化置于中心位置。近期的研究工作集中在使用 Inkubalm（由 Lelapa AI 开发）的小型语言模型上，该模型针对多种低资源非洲语言进行了优化，并在翻译和 NER [62] 上取得了最先进的结果。为了获得更全面的非洲语言建模工作的概览，请参阅此综述 [63]。

B. 部署和差距

非洲的几家保险公司已将其运营与 AI 和 LLM 技术相结合。这些公司包括**菠萝**，这是一个面向千禧一代的保险公司 [64]，它与 Old Mutual 等业界领导者合作

开发解决方案，以简化理赔处理流程。例如，他们允许客户上传图片进行车辆索赔评估。在肯尼亚，**普拉顾问公司**利用 AI 和遥感数据为跨越 12 个国家的超过 150 万小农户大规模提供农业保险 [65]。他们的平台使用卫星图像和预测性天气模型来实现自动赔付，并将 LLM 集成到其基于聊天的客户支持工具中。也在肯尼亚：**涡阿科鸟**向超过 350 万保单持有人提供移动健康和生活微保险 [66]。他们将大型语言模型集成到客户服务聊天机器人和理赔验证流程中，使得平均理赔周转时间低至 4 小时 [67]。

裸保险，一家总部位于南非的公司，建立了一个完全数字化的承保和理赔引擎。他们使用 GPT 系列模型来驱动投保人面对的聊天机器人，实现了大约 60%的端到端索赔完成率。据报道，他们可以在 15 分钟内解决简单案件，并将与索赔过程相关的运营成本降低了近 40% [68]。同样在南非，**莱拉帕人工智能**，由一组研究人员领导，提供了 Vulavula API [69]，包含五个模块：分析、对话、转录、翻译和讲话。这些模块处理非洲语言的音频、文本和对话任务。他们的模块化方法非常适合转变保险和金融中的不同任务的整体方式。人工智能的快速发展，加上数据量的增长、复杂性和可用性的提升，已经通过增强建模精度和效率改变了保险领域 [70]，允许精算师更快地建立模型，并让精算师优先处理依赖于人类判断的任务，这是目前无法完全通过没有人工干预的人工智能实现的。这些变化已经在精算师采用 LLMs 以丰富其工作流程中显现出来，合并结构化和非结构化数据，自动化重复过程，并提高风险模型的复杂性 [71]。此外，LLMs 的其他可能用途包括理解和提供监管报告的结果，例如 IFRS 17 和其他相关指南。通过将指南文档存储在一个本地搜索索引中，一个 LLM 代理可以在查询时查找它需要的确切部分，然后使用这些段落来构建其答案，并引用所参考的具体文件或章节，确保准确性和透明度。嵌入 IFRS 17 文本、数值文档和 PDF 到向量数据库中，使 LLM 代理能够按业务线自动化索赔和保费报告，以精确引用来回答监管查询，并为非技术用户生成报告和查询构建代码，提供一种实用的、自动化的和透明的合规方法 [72]。

C. 高影响应用的机会

初始 LLMs 在保险行业中将代理 AI 应用于文本，以分类身体伤害并从多语言事故报告中估计伤害程度 [73]，为精算师通过解释复杂的政策和索赔数据设计个性化产品奠定了基础 [72]。除了索赔处理 [74], [75]、承保和定价 [76], [77] 以及死亡率建模 [78] 之外，非洲的人工智能承诺基于远程信息处理使用和适应天气条件的辅助驾驶系统来制定政策 [79], [80]。这些有潜力重塑个性化的风险配置文件，并且可以潜在地整合到代理人工智能框架中。我们还设想预测性诊断 [81]、远程患者监控 [82] 以及个性化治疗计划 [83] 将通过个性化实时评估转变医疗保险，允许动态保费和保障范围。

IV. 政策、治理与合作路径

本节讨论政策、治理和合作对非洲保险业人工智能采纳方向和影响的作用。

A. 三重 R：规则、风险和现实

随着 LLMs 在各行业的广泛采用 [84]，对其适当整合的指南需求已经引起了个人和政府的广泛关注 [85]。虽然对 LLMs 的监管以某种形式存在，但其实施方式各异：例如，南非尚未通过专门的人工智能政策（截至撰写本文时）。然而，现有的法律，如《个人信息保护法》(POPIA) [86]，对处理个人可识别信息施加了限制，间接影响了 LLMs 的部署 [87]。南非和肯尼亚是非洲两大经济体 [88]，各自提出了人工智能监管框架，概述了未来目标和好处 [89], [90]。理解这些规则对于使用 LLMs 的企业至关重要。

这一监管活动的激增强调了公司，特别是那些在其工作流程中使用或探索 LLMs 的公司，需要仔细考虑处理潜在可识别信息 (PII) 的影响，鉴于其对就业安全和更广泛经济的潜在影响 [91]。许多公司已经采用 LLMs 来执行电子商务、教育和金融等领域的工作 [92]，因此认识到有必要适应并整合这些系统以惠及各行业的业务流程至关重要。随着行业继续将 LLMs 集成到诸如金融和保险等关键领域，越来越清楚的是，仅靠监管无法确保 AI 的安全、有效和公平采用 [93]。解决自动化决策、数据治理和系统互操作性等问题需要超越合规的协调方法。这强调了培养合作生态系统的重要性，在这种生态系统中，监管机构、保险公司、精算师和技术开发者共同设计符合伦理和行业特定现实的包容性和可持续 AI 部署。

B. 构建包容性生态系统

LLMs 和代理 AI 成功融入金融领域的直接和间接因素与不断变化的政策及监管框架紧密相关。例如，POPIA [86] 对如何训练和部署 AI 施加了具体的数据可用性、处理和存储方面的合规要求，影响了这一过程。因此，监管限制不仅塑造了 LLMs 的技术实施方式，还建立了系统故障或误用时的责任机制。最近 Meta [94]、OpenAI、Anthropic [95] 和 Google [96] 的诉讼案就体现了这一点，并为 AI 设定了未来的监管框架。将高级 AI 整合到严格的遗留系统中：这在高度受监管的金融和保险领域很常见，不仅需要提升运营效率，还需要确保公平性、透明度以及完全遵守法规 [97]。这种整合要求精算师、保险公司、技术开发者及监管机构之间的紧密合作。作为领域专家，精算师需要接受培训以量化由 AI 驱动决策的风险，特别是在高风险应用中 [98], [99]。技术团队还需要构建特别易于解释、伦理和适应不断变化的法规的模型。将这些目标嵌入到整体组织的关键绩效指标中并建立多学科团队确保它们在整个企业中的代表性和沟通。在监管方面，需要采取更加主动的方法来维护 [100]。通过与行业利益相关者早期接触，监管部门可以确保制定出实用有效的指导方针，这些指南能够解决特定行业的风险，同时不会因过于规避风险而阻碍创新。我们坚信：只有所有方参与其中，才能在保险领域提供既具创新性又在当地有效且公平的人工智能解决方案，这一观点得到了 [101] 的支持。

V. 结论与行动呼吁

总之，本研究探讨并揭示了将大型语言模型和代理人工智能整合到保险行业中的现有努力，特别是在非洲地区。我们强调这些技术如何重塑全球市场的工作流程、监管思维以及客户体验。我们展示了在快速发展的领域中经常被忽视的非洲贡献，并提出“主场优势”是我们的胜算所在：非洲的研究人员、保险公司、开发者、监管者及最重要的是被保险人，不仅有机会采纳全球解决方案，还能构建独特的地方性方案。我们提出了一个简单明了的行动号召：我们需要非洲人思考、创造并维护基于人工智能的保险解决方案在非洲市场的应用。显而易见的原因是：你的模型通常与你的数据一样好，而我们生成这些数据，生活在这个现实背后，比任何人都更拥有和理解它。

参考文献

[1] N. N. Mabaso, “Evaluation of the implementation of the twin peaks financial sector regulatory model in south africa,” Ph.D. dissertation, University of Pretoria, 2021. [Online]. Available: https://repository.up.ac.za/bitstreams/106d7444-3186-4884-ae40-53b858d8d611/download

[2] M. Seiranian and A. Suzuki, “Large language model parameter efficient fine-tuning for mathematical problem solving,” Ph.D. dissertation, 07 2024.

[3] E. Luciano, M. Cattaneo, and R. Kenett, “Adversarial ai in insurance: Pervasiveness and resilience,” *arXiv preprint arXiv:2301.07520*, 2023.

[4] Bank for International Settlements, “Regulating ai in the financial sector: Recent developments and main themes,” BIS FSI Insights Paper No. 63, 2025. [Online]. Available: https://www.bis.org/fsi/publ/insights63.pdf

[5] C. E. Shannon, “Prediction and entropy of printed english,” *The Bell System Technical Journal*, vol. 30, no. 1, pp. 50–64, 1951.

[6] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, p. 1735 – 1780, 1997.

[7] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” *arXiv preprint arXiv:1412.3555*, 2014.

[8] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, “Gpu computing,” *Proceedings of the IEEE*, vol. 96, no. 5, pp. 879–899, 2008.

[9] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, “Recurrent neural network based language model,” in *INTERSPEECH*, 2010, pp. 1045–1048.

[10] T. Brants and A. Franz, “Web 1t 5-gram version 1,” Linguistic Data Consortium, LDC2006T13, 2006.

[11] R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, and Y. Wu, “Exploring the limits of language modeling,” *arXiv preprint arXiv:1602.02410*, 2016. [Online]. Available: https://arxiv.org/abs/1602.02410

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, 🄵 Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008.

[13] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving language understanding by generative pre-training,” *arXiv preprint arXiv:1801.06146*, 2018.

[14] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language models are unsupervised multitask learners,” OpenAI blog, 2019, https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

[15] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 1877–1901.

[16] D. Hendrycks, C. Burns, S. Basart, A. Zou, M. Mazeika, D. Song, and J. Steinhardt, "Measuring massive multitask language understanding," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

[17] E. M. Bender, T. Gebru, A. McMillan-Major, and M. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* ACM, 2021, pp. 610–623.

[18] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[19] H. Touvron, L. Martin, K. Stone, P. Albert, M. Douze, N. Goyal, T. Raya, C. Raffel, and A. Joulin, "Llama: Open and efficient foundation language models," 2023.

[20] OpenAI, "GPT-4 technical report," https://cdn.openai.com/papers/gpt-4.pdf, 2024.

[21] Anthropic, "Introducing claude 2: A next-generation conversational llm," 2023, https://www.anthropic.com/product-overview/claude-2.

[22] DeepSeekR Team, "Deepseekr1: A multimodal retrieval-augmented foundation model," 2023, https://github.com/deepseekr/deepseekr1.

[23] N. Stiennon, L. Ouyang, J. Wu, D. M. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. Christiano, "Learning to summarize with human feedback," *arXiv preprint arXiv:2009.01325*, 2020.

[24] T. Lepikhin, H. Lee, Y. Xu, T. Chen, O. Firat, M. Huang, A. Jolicoeur-Martineau, and N. Shazeer, "Gshard: Scaling giant models with conditional computation and automatic sharding," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://arxiv.org/abs/2006.16668

[25] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022.

[26] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," *Advances in neural information processing systems*, vol. 33, pp. 5776–5788, 2020.

[27] S. Huang, D. Siddarth, L. Lovitt, T. I. Liao, E. Durmus, A. Tamkin, and D. Ganguli, "Collective constitutional ai: Aligning a language model with public input," in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2024, p. 1395 – 1417.

[28] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann, "Bloomberggpt: A large language model for finance," *arXiv preprint arXiv:2303.17564*, 2023.

[29] J. Ding, K. Feng, B. Lin, J. Cai, Q. Wang, Y. Xie, X. Zhang, Z. Wei, and W. Chen, "Insqabench: Benchmarking chinese insurance domain question answering with large language models," *arXiv preprint arXiv:2501.10943*, 2025.

[30] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, "Biogpt: Generative pre-trained transformer for biomedical text generation and mining," in *arXiv preprint arXiv:2210.10341*, 2022.

[31] D. Shu, H. Zhao, X. Liu, D. Demeter, M. Du, and Y. Zhang, "Lawllm: Law large language model for the us legal system," *arXiv preprint arXiv:2407.21065*, 2024.

[32] L. Martin, B. Muller, V. Suéry, É. d. l. C. Marchand, C. Soussen, and C. Darve, "Camembert: a tasty french language model," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, 2020, pp. 7203–7219.

[33] S. Cañete and J. Tiedemann, "Spanish-bert: A large-scale language model for spanish," 2021, https://github.com/dccuchile/beto.

[34] Y. Cui, W. Che, T. Liu, and G. Hu, "Chinese-bert-wwm: Whole word masking for chinese bert," 2019, https://github.com/ymcui/Chinese-BERT-wwm.

[35] B. Kossaï, M. Zeggour, J. Mpiana, and V. Umapathi, "Afriberta: A bert-based language model for african languages," in *Proceedings of the Second Workshop on Language Technologies for African Languages*, 2022, pp. 1–8.

[36] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, "ReAct: Synergizing reasoning and acting in language models," in *International Conference on Learning Representations (ICLR)*, 2023.

[37] J. Liu and B. Mozafari, "Query rewriting via large language models," *arXiv preprint arXiv:2403.09060*, 2024.

[38] H. Chase, "Langchain: A framework for developing applications with large language models," https://github.com/hwchase17/langchain, 2023.

[39] J. Kwok *et al.*, "Llamaindex (gpt index): A data framework for your llm application," https://github.com/run-llama/llama_index, 2023.

[40] C. Andersen, M. Shaddad, D. Lemaire *et al.*, "Microsoft semantic kernel: An orchestration sdk for large language models," https://github.com/microsoft/semantic-kernel, 2023.

[41] S. Salman *et al.*, "Auto-gpt: Autonomous ai agents powered by gpt-4 and beyond," https://github.com/Significant-Gravitas/Auto-GPT, 2023.

[42] P. Huryn, "The ultimate guide to ai agents for pms," The Product Compass Premium newsletter, jun 2025, accessed: 2025-07-03. [Online]. Available: https://www.productcompass.pm/p/ai-agents

[43] J. McCurry, "Japanese company replaces office workers with artificial intelligence," *The Guardian*, accessed: 2025-07-03. [Online]. Available: https://www.theguardian.com/technology/2017/jan/05/japanese-company-replaces-office-workers-artificial-intelligence-ai-fukoku-mutual-life-insurance?

[44] M. Orhan, "Building high performance document ocr systems," Tractable AI Blog, Jun. 2024, accessed: 2025-07-04. [Online]. Available: https://tractable.ai/building-high-performance-document-ocr-systems/

[45] Tractable, "Geico partners with tractable to accelerate accident recovery with ai," Press release, 2021, accessed: 2025-07-04. [Online]. Available: https://www.prnewswire.com/news-releases/geico-partners-with-tractable-to-accelerate-accident-recovery-with-ai-301299295.html

[46] A. Clere, "Tractable snaps up $65m led by softbank for car and property damage appraisals using ai," *TechCrunch*, accessed: 2025-07-04. [Online]. Available: https://techcrunch.com/2023/07/18/tractable-snaps-up-65m-led-by-softbank-for-car-and-property-damage-appraisals-using-ai

[47] A. C. Kumar, "Llm-powered chatbots for customer support in the insurance industry," *Aidetic Blog*, accessed: 2025-07-04. [Online]. Available: https://blog.aidetic.in/llm-powered-chatbots-for-customer-support-in-the-insurance-industry-7c80b0aa3163

[48] Shift Technology, "About shift technology," Company website, 2025, accessed: 2025-07-04. [Online]. Available: https://www.shift-technology.com/about

[49] Shift Technology Claims, "Ai-powered claims fraud detection for p&c insurers," Company website, 2025, accessed: 2025-07-04. [Online]. Available: https://www.shift-technology.com/resources/research/generative-ai-in-insurance-use-cases-examples-and-real-results

[50] PricewaterhouseCoopers, "Ai integration in legacy insurance systems: Challenges and solutions," PwC Global Report, 2024. [Online]. Available: https://www.pwc.com/ai-insurance-legacy-systems

[51] Deloitte, "Generative ai in insurance: Navigating costs and maintenance," Deloitte Insights, 2025. [Online]. Available: https://www.deloitte.com/insights/insurance-generative-ai

[52] L. Pretorius and A. Bosch, "Deepmorph: A deep learning approach to morphological analysis for south african bantu languages," *Computational Linguistics in Africa*, vol. 12, no. 1, pp. 45–60, 2020.

[53] V. Marivate, M. Mots' öehli, V. Wagner, R. Lastrucci, and I. Dzingirai, "Puoberta: Training and evaluation of a curated language model for setswana," in *Communications in Computer and Information Science, Artificial Intelligence Research*. Springer Nature Switzerland, 2023, pp. 253–266. [Online]. Available: https://arxiv.org/abs/2310.09141

[54] O. Oseni, N. Adegboye, and F. Lumor, "Building word embeddings for yoruba using deep learning," in *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC)*, 2020, pp. 3503–3510.

[55] C. Eze, I. Ekwueme, and C. Obi, "Igbobert: Building a pretrained language model for igbo," in *Proceedings of the 5th Workshop on African Language Technology*, 2022, pp. 45–54.

[56] Z. Abdullahi, I. Musa, and F. Bello, "Hausabert: A transformer-based pretrained model for hausa," in *Proceedings of the Workshop on African Language Technology*, 2021, pp. 23–32.

[57] K. Uwa, O. Nwachukwu, and P. Okoro, "Fine-tuning pretrained bert models for named-entity recognition and sentiment analysis in nigerian languages," in *Proceedings of the 4th Conference on African Language NLP*, 2023, pp. 112–121.

[58] J. Ngugi, S. Mburu, and E. Mutahi, "A bidirectional lstm approach to named-entity recognition in swahili," in *Proceedings of the 10th African Conference on Machine Learning and Data Mining*, 2019, pp. 120–131.

[59] D. I. Adelani, D. Babatunde, B. T. Gebre, M. Foufi, T. J. Humphreys, F. Nwachukwu, and D. Msihid, "Masakhane: Machine translation for african languages," in *Proceedings of the Third Conference on Machine Translation (WMT)*, 2021, pp. 497–503.

[60] J. Ojo, K. Ogueji, P. Stenetorp, and D. I. Adelani, "How good are large language models on african languages?" *arXiv preprint arXiv:2311.07978*, 2023.

[61] L. DeWitt Prat, O. Ndlovu Lucas, C. Golias, and M. Lewis, "Decolonizing llms: An ethnographic framework for ai in african contexts," in *Ethnographic Praxis in Industry Conference Proceedings*, no. 1, 2024, pp. 46–85.

[62] A. L. Tonja, B. F. Dossou, J. Ojo, J. Rajab, F. Thior, E. P. Wairagala, A. Aremu, P. Moiloa, J. Abbott, V. Marivate *et al.*, "Inkubalm: A small language model for low-resource african languages," *arXiv preprint arXiv:2408.17024*, 2024.

[63] J. O. Alabi, M. A. Hedderich, D. I. Adelani, and D. Klakow, "Charting the landscape of african nlp: Mapping progress and shaping the road ahead," *arXiv preprint arXiv:2505.21315*, 2025.

[64] Pineapple. (2024) The future of car insurance in south africa: How ai is revolutionising the industry. Pineapple Insurance. Blog post. [Online]. Available: https://pineapple.co.za/post/the-future-of-car-insurance-in-south-africa-how-ai-is-revolutionising-the-industry/

[65] A. Khanna, "Kenyan insurtech startup pula raises $6m series a," TechCrunch, accessed: 2025-07-04. [Online]. Available: https://techcrunch.com/2021/01/25/kenyan-insurtech-startup-pula-raises-6m-series-a-led-by-tlcom-capital/

[66] Turaco Insurance, "About us," Company website, 2025, accessed: 2025-07-04. [Online]. Available: https://www.turaco.insure/

[67] F. Company, "Turaco is making insurance more affordable to millions in africa," *Fast Company*, accessed: 2025-07-04. [Online]. Available: https://www.fastcompany.com/91037559/turaco-most-innovative-companies-2024

[68] Naked Insurance3, "Ai-driven claims automation at naked," Naked Insurance blog, 2023, accessed: 2025-07-04. [Online]. Available: https://naked.insure/blog/ai-driven-claims-automation

[69] A. Tsanni, "This company is building ai for african languages," MIT Technology Review, 2023, accessed: 2025-07-04. [Online]. Available: https://www.technologyreview.com/2023/11/17/1083637/lelapa-ai-african-languages-vulavula/

[70] R. Richman, "Ai in actuarial science," insureAI / University of the Witwatersrand, SSRN Electronic Journal SSRN 3218082, July 2018. [Online]. Available: https://ssrn.com/abstract=3218082

[71] C. Balona, "Operationalizing llms: A guide for actuaries," *Actuarial Innovation and Technology Strategic Research Program Technical Report*, 2025. [Online]. Available: https://www.soa.org/48ffb2/globalassets/assets/files/resources/research-report/2025/oper-genai-act-report.pdf

[72] ——, "Actuarygpt: Applications of large language models to insurance and actuarial work," *British Actuarial Journal*, vol. 29, p. e15, 2024.

[73] A. Troxler and J. Schelldorfer, "Actuarial applications of natural language processing using transformers: Case studies for using text features in an actuarial context," *British Actuarial Journal*, vol. 29, p. e4, 2024.

[74] A. Dimri, S. Yerramilli, P. Lee, S. Afra, and A. Jakubowski, "Enhancing claims handling processes with insurance based language models," in *2019 18th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2019, pp. 1750–1755.

[75] P. Jani and S. Anand, "Compliance-aware ai adjudication using llms in claims engines (delta lake+ langchain)," *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, vol. 5, no. 2, pp. 37–46, 2024.

[76] E. Krasheninnikova, J. García, R. Maestre, and F. Fernández, "Reinforcement learning for pricing strategy optimization in the insurance industry," *Engineering applications of artificial intelligence*, vol. 80, pp. 8–19, 2019.

[77] R. Balasubramanian, A. Libarikian, and D. McElhaney, "Insurance 2030 - the impact of ai on the future of insurance," *McKinsey & Company*, pp. 1–10, 2018.

[78] P. Chung, C. T. Fong, A. M. Walters, N. Aghaeepour, M. Yetisgen, and V. N. O' Reilly-Shah, "Large language model capabilities in perioperative risk prediction and prognostication," *JAMA surgery*, vol. 159, no. 8, pp. 928–937, 2024.

[79] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167–181, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0965856415000804

[80] M. Mots'oehli, F. Chen, H. W. Chan, I. Tlali, T. Babeli, K. Baek, and H. Chen, "Simulating refractive distortions and weather-induced artifacts for resource-constrained autonomous perception," arXiv preprint arXiv:2507.05536, 2025, accepted to the ICCV 2025 Workshop on Computer Vision for Developing Countries (CV4DC). [Online]. Available: https://arxiv.org/abs/2507.05536

[81] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, 2017.

[82] E. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.* New York: Basic Books, 2019.

[83] S. Reddy, J. Fox, and M. P. Purohit, "Artificial intelligence-enabled healthcare delivery," *Journal of the Royal Society of Medicine*, vol. 112, no. 1, pp. 22–28, 2019.

[84] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[85] OpenAI, "GPT-4 technical report," https://cdn.openai.com/papers/gpt-4.pdf, 2024.

[86] Information Regulator (South Africa), "Protection of personal information act (popi act)," https://popia.co.za/, n.d.

[87] White & Case LLP, "Ai watch: Global regulatory tracker - south africa," https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-south-africa, 2024.

[88] Statista, "Gdp of african countries by country," https://www.statista.com/statistics/1120999/gdp-of-african-countries-by-country/, 2024.

[89] Ministry of Information, Communications and the Digital Economy (Kenya), "Kenya artificial intelligence strategy 2025 – 2030," https://ict.go.ke/sites/default/files/2025-03/Kenya%20AI%20Strategy%202025%20-%202030.pdf, 2025.

[90] Pinsent Masons, "South africa takes first steps towards ai regulation," https://www.pinsentmasons.com/out-law/news/south-africa-takes-first-steps-towards-ai-regulation, 2024.

[91] A. Jaldi, "Artificial intelligence revolution in africa: Economic opportunities and legal challenges," *Policy Cent. New South*, pp. 2023–07, 2023.

[92] M. Raza, Z. Jahangir, M. B. Riaz, M. J. Saeed, M. A. Sattar *et al.*, "Industrial applications of large language models," *Scientific Reports*, vol. 15, 2025, art. no. 13755.

[93] A. Morales-Forero, S. Bassetto, and E. Coatanea, "Toward safe ai," *AI & SOCIETY*, vol. 38, no. 2, pp. 685–696, 2023.

[94] C. Reed, "Meta fends off authors' us copyright lawsuit over ai," https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/, 2025, accessed: 2025-07-04.

[95] V. News, "Judge backs anthropic: Ai training on legal books ruled fair use," https://www.vktr.com/ai-technology/judge-backs-anthropic-ai-training-on-legal-books-ruled-fair-use/, 2025, accessed: 2025-07-04.

[96] R. Staff, "Google's ai overviews hit by eu antitrust complaint from independent publishers," https://www.reuters.com/legal/litigation/googles-ai-overviews-hit-by-eu-antitrust-complaint-independent-publishers-2025-07-04/, 2025, accessed: 2025-07-04.

[97] B. Gapp, F de Fauconberg, A. Bessis, H. Goble, and N. Obodoekwe, "Digital finance in africa: Accelerating foundations for inclusive and sustainable local innovation," *Environmental Sciences Proceedings*, vol. 15, no. 1, 2022. [Online]. Available: https://www.mdpi.com/2673-4931/15/1/66

[98] V. Galaz, M. A. Centeno, P. W. Callahan, A. Causevic, T. Patterson, I. Brass, S. Baum, D. Farber, J. Fischer, D. Garcia, T. McPhearson, D. Jimenez, B. King, P. Larcey, and K. Levy, "Artificial intelligence, systemic risks, and sustainability," *Technology in Society*, vol. 67, p. 101741, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0160791X21002165

[99] J. Göpfert, J. M. Weinand, P. Kuckertz, N. Pflugradt, and J. Linßen, "Risks of ai-driven product development and strategies for their mitigation," 2025. [Online]. Available: https://arxiv.org/abs/2506.00047

[100] M. Fenwick, E. P. Vermeulen, and M. Corrales, "Business and regulatory responses to artificial intelligence: Dynamic regulation, innovation ecosystems and the strategic management of disruptive technology," *Robotics, AI and the Future of Law*, pp. 81–103, 2018.

[101] P. Graça and L. M. Camarinha-Matos, "Assessment of sustainable collaboration in collaborative business ecosystems," *Computers*, vol. 10, no. 12, p. 167, 2021.