arXiv:2304.03748v1 [cs.LG] 7 Apr 2023

地球系统人工智能架构与协同设计的视角

可预测性

马鲁蒂·K.·穆杜努鲁,[1] 詹姆斯·A.·昂,[1] 马汉特什·哈拉潘纳瓦尔,[1] 西蒙·D.·哈蒙德,[2] 玛雅·B.·戈卡莱,[3] 詹姆斯·C.·何,[4] 图沙尔·克里希纳,[5] 萨拉特·S.·斯里帕蒂,[6] 马修·R.·诺曼,[6] 艾维·B.·彭,[7] 和菲利普·W.·琼斯[8]

[1]太平洋西北国家实验室，美国华盛顿州里奇兰。[2]国家核安全管理局（*NNSA*），美国哥伦比亚特区华盛顿。[3]劳伦斯利弗莫尔国家实验室，美国加利福尼亚州利弗莫尔。[4]卡内基梅隆大学，美国宾夕法尼亚州匹兹堡。[5]佐治亚理工学院，美国佐治亚州亚特兰大。[6]橡树岭国家实验室，美国田纳西州橡树岭。[7]瑞典皇家理工学院，瑞典。[8]洛斯阿拉莫斯国家实验室，美国新墨西哥州洛斯阿拉莫斯。

通讯作者: M. K. Mudunuru，电子邮箱：maruti@pnnl.gov；太平洋西北国家实验室-SA-183269；LA-UR- 23-22945；LLNL-JRNL-846885

摘要：近期，美国能源部（DOE）科学办公室（Office of Science）下属的生物与环境研究（BER）和先进科学计算研究（ASCR）项目组织并举办了地球系统可预测性人工智能（AI4ESP）系列研讨会。研讨会得出的一项重要结论是，美国能源部 BER 和 ASCR 社群认为，有必要为地球系统可预测性开发一种新范式，推动人工智能（AI）贯穿野外、实验室、建模和分析等各项活动，这一范式称为 ModEx。BER 的“模型—实验”（ModEx）是一种迭代方法，可利用过程模型生成假设。形成的假设为野外和实验室工作提供依据，以采集测量和观测数据；随后利用这些数据对模型（例如基于过程的模型）预测进行参数化、驱动和检验。AI4ESP 系列研讨会共举办了 17 场技术专题会议。本文探讨“人工智能架构与协同设计”专题会议及其相关成果。该专题会议包括两场特邀报告、两场全体讨论会，以及三个分组讨论室，分别探讨以下具体主题：（1）美国能源部高性能计算系统；（2）云端高性能计算系统；（3）边缘计算和物联网（IoT）。我们还就该协同设计领域的潜在研究提出前瞻性思路和观点，这些研究可通过与其他 16 个专题的协同作用得以实现。相关思路包括：（1）重新构想协同设计；（2）从数据采集到分发；（3）异构高性能计算解决方案，以整合人工智能/机器学习以及不确定性量化等其他数据分析方法与地球系统模拟和仿真；（4）将人工智能赋能的传感器整合到地球系统测量和观测中。这些观点正是本文的鲜明特色。

重要性声明：本研究旨在为地球系统可预测性的人工智能架构和协同设计方法提供观点。这些前瞻性观点至关重要，因为人工智能赋能的模型—数据集成已显示出提升与气候变化、扰动和极端事件相关预测能力的潜力。我们的前瞻性理念为协同设计的下一步发展提供指导，利用最先进及未来的计算基础设施，增强地球系统模型、观测和理论。

1. 引言

美国能源部（DOE）近期结束了地球系统可预测性人工智能（AI4ESP）研讨会（Hickmon 等，2022）。此次研讨会由美国能源部科学办公室下属的生物与环境研究（BER）项目和先进科学计算研究（ASCR）项目主办。来自世界各地的研究人员共参与了17场会议，讨论人工智能（AI）如何在实地、实验室、建模和分析活动中提升整个领域的地球系统可预测性（Hoffman 等，2017，图1.3）。讨论的主要重点是利用人工智能转变BER的“模型—实验”（ModEx）集成（Chambers 等，2012，第93页）。

传统上，ModEx范式（Hoffman等，2017年，第1节）以迭代方式，将在野外或实验室开展的观测、实验和测量与概念模型/过程模型相结合。近年来，人工智能的发展显示出高效加速传统ModEx的潜力（Tsai等，2021年；Cromwell等，2021年；Mudunuru等，2022年）。ModEx循环中的这种转变，对于高效且准确地整合美国能源部的观测能力和平台1、过程模型与软件基础设施2以及计算硬件[3]至关重要。然而，要实现这一人工智能赋能的ModEx愿景，还需要在协同设计及相关人工智能架构方面取得重大进展（Germann等，2013年；Zhang等，2019年；Beckman等，2020年；Descour等，2021年；Bringmann等，2021年）。本文提出了关于人工智能架构和协同设计方法的观点，这些架构和方法对于开发面向地球系统可预测性的人工智能赋能ModEx至关重要。这些观点包括对计算和存储

[1]常见的 BER 观测能力包括大气辐射测量气候研究设施（ARM）（ARM 2022）和环境分子科学实验室（EMSL）（EMSL 2022）
[2]由美国能源部资助的最先进开源大规模并行多物理场代码包括 PFLOTRAN（Lichtner 等，2020）、高级陆地模拟器（ATS）（ATS 2022）和能源部百亿亿次级地球系统模型（E3SM）（E3SM 2022）[3]由美国能源部科学办公室高级科学计算研究计划资助的计算基础设施和科学用户设施包括阿贡领导力计算设施（ALCF）（ALCF 2022）、美国国家能源研究科学计算中心（NERSC）（NERSC 2022）和橡树岭领导力计算设施（OLCF）（OLCF 2022）

基础设施进行协同设计，以实现自动化机器学习特征工程和模型选择；整合传感器、过程模型与机器学习方法，以实现高效的数据同化。我们还提出了关于协同设计框架和平台的前瞻性系统构想，以帮助BER社区加速人工智能架构在ModEx生命周期中的应用。

本文的概要如下：第 2 节介绍了 AI4ESP 研讨会参与者讨论的人工智能架构与协同设计领域的最新进展。第 3 节提出四种不同的未来构想，第 4 节探讨了实现这些构想所面临的重大挑战。我们还讨论了克服这些重大挑战的近期、中期和长期目标。第 5 节阐述了潜在研究方向，以期与 AI4ESP 研讨会的其他环节形成协同效应。第 6 节给出结论。

2. 科学现状

本节介绍人工智能架构和协同设计的科学技术现状。重点是用于捕获和整理数据、开发先进的人工智能/机器学习模型，以及进行推理以量化和提升地球系统建模与模拟可预测性的计算资源和美国能源部用户设施。

a. 美国能源部的高性能计算用户设施

过去几十年来，美国能源部已投入数亿美元，用于开发高性能计算（HPC）用户设施（Stevens 等，2020；Vetter等，2022；Heroux 等，2022）。美国能源部面向百亿亿次级计算的投资包括阿贡国家实验室的阿贡领导力计算设施（ALCF）（例如 Aurora）、橡树岭国家实验室的橡树岭领导力计算设施（OLCF）（例如 Frontier）以及美国国家能源研究科学计算中心（NERSC）（例如 Perlmutter）。在 2022年 11月的 Top500 榜单中，Frontier被评为速度最快的超级计算系统。美国能源部领导力级计算设施的最新一代产品，建立在将中央处理器（CPU）和图形处理器（GPU）集成到异构系统中的基础上。与此同时，美国能源部的生物与环境研究计划已在先进科学模型方面投入大量资源（E3SM 2022；Lichtner等，2020；ATS 2022），其中包括旗舰级能源部百亿亿次级地球系统模型（E3SM）（E3SM 2022），该模型专门设计用于实现对百亿亿次级超级计算机的高效利用。

这些高性能计算资源显著提升了多个领域模型的可预测性，包括地球系统模拟、地下流动与输运模型等（例如 E3SM、PFLOTRAN）。作为美国能源部百亿亿次计算项目的一部分，精选的一组地球科学应用——E3SM-MMF（2022）和 Subsurface-ECP（2022）——坚定地聚焦于面向百亿亿次计算时代的模型开发。此外，ECP 旗下的 E3SM-MMF 子项目（E3SM-MMF 2022）等工作开展了有针对性的协同设计活动，包括与供应商密切合作，及早评估架构并设计算法。这些工作的经验表明，有必要将协同设计拓展至人工智能架构，并扩大地球科学应用的覆盖范围。这些进展使我们能够在利用 GPU 人工智能/机器学习性能的商业驱动力的同时，实现 GPU 上的高能效性能。

随着摩尔定律放缓，计算领域认识到对架构专用化的需求日益增长。因此，下一代高性能计算系统很可能会在当前 CPU 与 GPU 混合设计的基础上进一步提升异质性。美国能源部在人工智能促进科学方面的工作（Baker 等，2019年；Stevens 等，2020年）正在探索相关能力，为高性能计算应用（例如阿贡领导力计算设施的人工智能测试平台（Testbed 2022））与数据科学及人工智能/机器学习框架的集成奠定基础。

b. 云计算

云服务提供商[4] 提供了易于使用的工具来运行人工智能/机器学习工作负载。然而，不同提供商的人工智能/机器学习工具功能和用户界面需要具备更好的兼容性，否则难以在云联合体中实现互操作性（Chouhan 等，2020；Rosa 等，2021；Saxena 等，2021）。目前，特定地球系统模型（ESM）的数据存储在云存储系统中（Xu 等，2019），但这些数据存储分散在各个团队和项目中，缺乏联合视图。云服务提供商目前能够提供 PB 级至 EB 级的数据存储容量。商业云的费用取决于数据访问、计算或分析。如果频繁向云端传输数据或从云端传出数据，费用可能会高得令人难以承受。商业人工智能/机器学习云基础设施和服务主要面向文本和图像数据。云服务提供商已展示了面向这些应用的大规模人工智能能力。例如，参数量接近 1 万亿的最大型人工智能自然语言处理（NLP）模型，已在 Selene 上实现（Chen 等，

[4]常见提供商包括亚马逊云科技（AWS）（AWS 2022）、Google Cloud Platform（GCP）（GCP 2022）和 Azure（Azure 2022）

2019年）（该系统在 2022 年 11 月的 Top 500 榜单中位列第九快的超级计算系统）。云端已有针对特定应用的工作流服务，其中包括许多人工智能/机器学习方法；云平台也提供了可用于创建更复杂工作流的基础组件。然而，目前尚无能够高效、准确地整合外部数据源或与高性能计算模拟协同运行的地球系统模型工作流。要以适合领域科学家的形式创建此类工作流，需要计算机科学专业知识（Chen 等，2017；Bauer 等，2021）。

c. 边缘计算

最近，用于模式分类、异常检测、数据压缩的无监督学习、边缘推理，以及利用流式传感器数据进行持续学习的人工智能方法，已在地球系统模型社区中得到广泛应用（Beckman et al. 2020；Talsma et al. 2022）。这一进展得益于人工智能模型在各类智能计算设备上的快速部署，这些设备包括连接传感器的 Raspberry Pi/Shake、Nvidia Jetson Nano、Google Coral Dev Board 和 Intel Neural Compute Stick（Catlett et al. 2017, 2020；Mudunuru et al. 2021）。边缘计算与智能传感器（例如 AI@SensorEdge）的融合，带来了许多不同的部署场景，包括在太空或专用地面站对美国国家海洋和大气管理局（NOAA）和美国国家航空航天局（NASA）的地球观测卫星图像进行边缘处理，或用于控制美国能源部（DOE）的辐射测量（ARM）或环境分子科学实验室（EMSL）用户设施仪器（Beckman et al. 2020）。我们还可以将边缘计算与各类分布式传感器结合，这些传感器为美国能源部（DOE）的 ARM 用户设施采集观测数据和测量数据。具备嵌入式硬件加速器的自适应传感器现已开始涌现（例如 Wagg-gle、PurpleAir）（Beckman et al. 2016；Stavroulas et al. 2020；Barkjohn et al. 2021）。分布式应用的新概念也在开发中，例如地貌计算，其中天气研究与预报模型以分布式、联邦式的方式部署，并能够根据环境动态调整（Daepp et al. 2022）。

3. Future System Concepts

在本节中，我们介绍了几种可能的未来系统构想，这些构想由参加AI4ESP 研讨会分组讨论会的参与者进行了讨论。讨论重点是演变

美国能源部领导的高性能计算和人工智能领导计算设施（LCF）系统。这些大规模异构计算系统为利用高性能计算推进人工智能架构和协同设计奠定了基础。此外，这些未来概念有望为未来地球系统建模和人工智能赋能的 ModEx 提供截然不同的方法。

a. 集中式大规模高性能计算概念

基线系统概念是大规模高性能计算和云计算系统的未来演进方向。下一步将推动后百亿亿次级架构超越第一代集成 CPU 和 GPU 的美国能源部异构系统。随着高性能计算和云计算社区日益依赖硬件专用化来提升性能，协同设计方法将支持开发加速器（Lie 2021；Reuther et al. 2021；Cortés et al. 2021），用于科学建模和人工智能/机器学习方法中频繁使用的内核。新型专用加速器还可能用于支持其他数据科学能力，例如不确定性量化、流式分析或图分析（Halappanavar et al. 2021；Acer et al. 2021）。这些未来的超高异构性大规模计算系统必须通过协同设计来支持与地球科学可预测性和科学机器推理相关的不断增长的计算规模和数据集规模（Yang et al. 2016；Zhang et al. 2020；Yu et al. 2022）。

b. 采用集中式高性能计算/云资源的边缘传感器概念

在第二种系统概念中，环境数据由遍布全球的大量点式传感器（Christensen and BlancoChia2017；Winter et al. 2021）和分布式传感器（例如光纤）（Lindsey et al.2019年）记录。这些先进的传感器旨在监测特定的关注对象（例如河流流量、营养物质、温度、化学浓度、光），并将这些数据传回一个集中式位置（Beckmanet al. 2020）。在这一集中式设施中，大型高性能计算或云计算环境将处理传入的数据流，以便将其整合到极端天气事件、气候、水文学及其对地球系统影响的在线模拟中。

我们可以在这一系统构想的多个环节中运用人工智能/机器学习能力。首先，传入系统的传感器数据速度可能会超出即使是最强大的数据处理中心的处理能力。因此，如此庞大的数据量不太可能被存储下来。

存储在内存中，或甚至存储在临时存储资源中（例如文件系统或对象存储）。先进的人工智能/机器学习模型可以经过训练和定制，用于汇总传入数据流或从中选取相关特征。此类编码或特征选择过程将大幅减少需要保留并整合到持续运行的模拟中的数据量。另一种可能性是，利用人工智能/机器学习模型从传入数据流中识别异常或前兆（Yuan 等人（2019）），据此判断模拟应重点关注的区域——例如飓风的形成，或发生重大雨落雪事件或野火的可能性很高的区域。

由于传感器可能需要部署或运行于分散且恶劣的环境中（例如偏远地点、极端温度或压力环境），对于所有可能的输入而言，可靠的数据流都不太可能传输到集中式位置。一个常见的应用场景是用于研究城市科学的智慧城市场景。图1是对各种已部署的传感器、计算和数据存储能力的概念性描绘（Zhu等，2021）。在这样的环境中，可以使用人工智能/机器学习模型来填补测量空白，并向未来在大规模计算资源上运行的模拟提供更加一致的观测数据视图。

c. 从边缘到数据中心的联邦处理概念

第三种潜在的系统设计通过在分布式传感器网络中或其附近利用更多的处理能力，扩展了第二种概念。我们可以直接在传感器本身或附近的边缘服务器（例如雾计算）上处理传感器数据，处理元件可以向其中传输少量传感器数据（Stevens等，2020，第15章）。随后，本地处理站可以将其原始数据或本地处理后的数据发送至集中的高性能计算和/或云资源，以便纳入模拟模型和集中的人工智能/机器学习模型中，如第一种系统概念所述。

这种方法的优势在于，可以在本地进行数据筛选和特征提取，大幅减少必须传输到集中式资源的数据量。假设能够在传感器之间建立性能足够出色的本地网络，那么过程模型参数、部分结果，甚至人工智能/机器学习模型更新，都可以在本地范围内交换，从而实现真正的联邦式设计。最初，这一概念利用现有的网关和为现场传感器提供服务的局域网。通过协同设计合作，可以将这项服务扩展到包括应用/传感器专用



图 1. 智慧城市场景中设有大量固定传感器部署点，用于测量温度、风廓线、CO_2 浓度、降水量等；此外，各种移动设备也可间歇性地用于补充测量和观测数据的采集。城市环境将支持 5G 乃至未来的 6G 等先进无线通信技术，以了解城市与气候之间的相互作用。 [图由太平洋西北国家实验室先进无线通信实验室制作]

处理，以筛选、分析、压缩、加密并统一通过无线网络传输测量数据的多个传感器流。

d. 动态自适应联邦处理概念

最后一种系统构想建立在前面三种构想的基础上，通过增强传感器本地资源分布式网络中的反馈与控制路径来实现这一目标（Di Lorenzo 等，2021；Charles 等，2021）。本地控制能够以更低的延迟做出决策，动态控制传感器网络观测、测量、记录和传送哪些信息（Morell 和 Alba，2022）。这种设计具有深远影响——通过在线动态控制传感器，地球天气和气候模拟实际上可以将传感器输入集中于特定的关注量或地理位置。例如，可以关注预计将发生极端天气事件的区域，或关注气候科学家确定需要哪些特定信息来帮助提高模型质量。

这一构想可扩展至多个高性能计算和/或云数据中心，以开展联邦式人工智能/机器学习建模。人工智能/机器学习模型可以在这一系统中发挥关键作用，持续自主地在线检查不断演进的模拟或已记录数据，识别数据不足或统计薄弱的区域。此外，动态自适应系统或许能够审慎地获取和筛选数据，以提升训练质量，从而减少长时间收集海量且可能难以处理的数据集的需求（Catlett 等，2017）。

4. 重大挑战

集成联邦处理的系统概念已超出当今可负担技术的能力范围。这将需要在基础技术系统和协同设计项目方面进行大量投资。气候科学家、数学家、人工智能/机器学习专家、计算机科学家和硬件工程师之间需要形成这种协同，以平衡人工智能赋能的 ModEx 所面临的性能、能耗、成本和安全性等相互竞争的挑战。以下小节将介绍以下领域中出现的技术挑战：（1）可编程性和易用性；（2）数据移动；（3）能效；以及（4）数据隐私与安全。

a. 可编程性和易用性

当前及近期的挑战是将科学建模与仿真应用同人工智能/机器学习方法相集成。这推动了将以 C/C++ 和/或 Fortran 编写的地球系统高性能计算应用，与使用基于 Python 的机器学习框架的人工智能/机器学习方法相集成的需求（Ott 等，2020 年）。目前正在开发编程模型，以支持应用和工作流向异构计算系统的融合。许多人工智能/机器学习架构都提供对低精度或混合精度的硬件支持，因此需要相应工具来分析具体模型组件中哪些部分可以利用这些能力。我们必须创建用于 ESP 数据共享和云端数据联邦的协议与工具。易用性方面的挑战在于，如何管理将融合后的应用工作负载映射到未来异构计算架构的复杂性；这些架构将专用硬件加速器与通用 CPU/GPU/张量处理单元处理器相集成。

领域科学家有兴趣探索新型异构先进架构计算系统的能力。通过在边缘端与传感器和人工智能分析进行接口连接，领域科学家将能够提取改进扰动建模所需的可执行信息。

和极端事件。大多数 ESP 应用都需要这种协同设计。例如，流域科学、水文学、生态水文学、气候变率与极端事件、气溶胶与云以及大气建模，都是 AISensorEdge 影响力最大的跨领域主题。与分布式传感器网络对接的协同设计方法将使我们能够：(1) 在扰动期间收集可靠且相关的流域数据；(2) 通过在大气辐射测量（ARM）仪器中嵌入智能，监测陆地—大气—沿海相互作用；(3) 近实时地了解野火事件及其对生态系统的影响；(4) 评估受极端事件影响的关键基础设施（例如，参见《AI4ESP 报告》第 9 章“人类系统与动态”）。

不过，要理解如何将 AI4ESP 工作流映射到多样化的计算系统选项集合，仍面临诸多挑战。理解最初为通用商业工作负载开发的人工智能/机器学习能力是否适用于地球系统可预测性（ESP）混合建模应用，或适用于观测与测量能力，至关重要。从集中式大规模建模与训练，到边缘计算推理和联邦学习，应用、算法与方法的组合和分布都带来了新的挑战。这为 AI4ESP 社区开发新一代用于建模和观测能力的代理应用与基准测试提供了重要机遇。例如，人工智能赋能的协同设计将使我们能够在传感器边缘模拟并部署美国能源部的代码（如 PFLOTRAN、ATS 和 E3SM），从而增强 ARM 仪器和 EMSL 用户设施的能力。重点应放在促进与硬件设计人员、系统软件开发人员、算法开发人员和领域科学家的沟通与协同设计合作上。

b. 数据移动

与完整、协调的地球传感器能力相关的预期数据量将是前所未有的。这类网络不仅会生成数量和多样性都达到以往难以想象程度的数据，对如此大规模数据进行处理、传输以及后续存储所带来的计算和网络负载，也将比当前任何可用系统高出几个数量级。就能耗和延迟而言，数据移动成本促使人们关注在 AI4ESP 科学生态系统中实现计算资源的联邦化与分布式部署。人工智能/机器学习技术可以通过识别模式和异常，以及对数据子集进行摘要，帮助减少这类数据量。我们

将需要在人工智能/机器学习方法上进行大量投入，以确保建模能力与所记录数据的类型相兼容并且高效，尤其是在这些数据可能偏离商用照片或视频能力的情况下。可能有助于实现高能效数据传输的技术包括：投资于硅光子网络能力，以及类似 5G 或 6G 的广域通信网络，使传感器无需实体布线即可在短距离或中距离范围内进行通信（Beckman 等，2020 年）。在存储方面，云技术（例如高性能、大容量的数据对象存储）很可能能够应对不断增加的稀疏数据存储量。然而，按照当前商用云服务的定价，这将构成一项巨大的成本障碍。我们还可以利用人工智能/机器学习为地球系统数据实现创新的压缩技术，在不增加存储成本的情况下提高信息密度。此外，美国能源部的高性能计算中心可以将云存储系统中的理念和方法融入未来的并行文件系统和存储系统，逐步实现此类能力。

c. 能源效率

集成传感器、联邦处理以及用于处理数据传输的广域通信网络构成的大规模网络，能耗可能会非常高。虽然这并非百亿亿次级计算的优先关注重点，但数据处理和通信仍然十分耗电。通过采用新型材料、设备和处理技术（例如用于分析图像/视频的类脑加速器），协同设计有望改善这一状况。不过，要实现大规模、高能效的传感网络，仍需在基础技术方面进行大量投入。通过协同设计平衡性能与能效，还将解决如何在地球系统模型科学生态系统中分配建模、机器学习、不确定性量化及其他流式分析能力的问题。通过协同设计，将美国能源部的异构高性能计算系统与云计算、边缘服务器以及配备物联网设备的传感器整合起来，将改变 ModEx 循环。

d. 数据隐私与安全

随着地球系统建模日益与分布式观测网络及可能的联邦式处理能力相结合，此类传感器网络中信息的质量、准确性和稳健性将变得愈发重要。如果

建模和测量能力所生成的信息用于支持影响重大的国家或国际科学政策决策，那么信息安全也必须得到保障。数据可能遭到篡改或恶意修改，其后果显而易见：作为准确科学预测依据的国家或国际资源可能会受到严重影响。在数据采集系统中，尤其需要重视数据隐私问题，因为系统可能会采集个人受试者的图像或视频，或从数据中识别其行为。例如，传感器可能具备识别人体系统数据模式的能力（如公民科学或城市环境中的相关数据）。协同设计在这一领域大有可为——从一开始就让安全专家参与信息物理系统设计，便能将安全的数据传输和处理作为优先要素加以集成，而非最后才作为软件衍生的附加层加入。此外，如果能将与特定个人相关的局部特征汇总到更大的联邦模型中，并在完整的系统模型中对个人模式进行模糊处理或删去，就有助于保护数据隐私。

5. 与其他 AI4ESP 研讨会环节的协同效应

在本节中，我们将从愿景角度探讨未来理念，以及与其他研讨会环节协同开展的潜在研究。表 1总结了这种协同关系，涵盖短期（< 5 年）、中期（5年）和长期（10年）目标。重点在于探讨人工智能架构和协同设计方法如何与综合水循环及相关水循环极端事件相互关联。

表1。本表列出了克服第4节所讨论重大挑战所需的短期、中期和长期目标。逐步实现这些具体目标，将使我们能够推进未来系统概念的发展，从而提高地球系统的可预测性。

短期目标	中期目标	长期目标	协同设计机遇
基准数据集	联邦学习的数据格式	提升 ESP 各领域的效率	极端事件异常分析
分布式人工智能/机器学习工作流程	用于不确定性量化的人工智能/机器学习	大规模人工智能演示	用于降尺度和升尺度的人工智能
人工智能/机器学习代理模型	人工智能/机器学习 + 物理模拟器	用于流式分析的人工智能	人工智能/机器学习 + 物联网 + 百亿亿次级生态系统
面向边缘的人工智能/机器学习抽象	传感器边缘人工智能	人工智能赋能的自动化	ESP 数字孪生

大气建模 – 需要推进跨尺度的网格下物理建模，并指导或自动化过程模型校准。这包括：(1) 用于参数化和跨尺度知识迁移的协同设计方法，以及 (2) 用于数据集、软件、测试、验证和训练工作流的人工智能基础设施，以实现高效的模型校准。

陆地建模 – 用于在陆地与大气模型之间高效传递信息的人工智能架构。这包括：(1) 用于捕捉网格内全部复杂性的子网格参数化，(2) 利用 LCF 捕捉异质性，以及 (3) 使用先进的人工智能架构（例如 Transformer模型）应对观测缺口。

水文学 – 需要先进的人工智能架构来进行参数估计、降尺度和插补，以改进数据产品。需要采用模型—数据协同设计方法，以确定在没有实际测量数据的情况下，需要多少观测数据以及何种类型的观测数据，才能达到预期的过程模型性能。这包括利用 5G 或其他高速网络及软件流水线，加快野外仪器与过程模型之间的信息传输，从而支持近实时采样决策。

流域科学 – 需要采用协同设计方法，以更好地了解：(1) 所收集数据的质量；(2) 在扰动和长期扰动作用下，利用基于过程的模型（例如 PFLOTRAN）预测流域响应（例如微生物活动的演变）的能力；(3) 何时、如何以及在何处收集数据（例如，野火、洪水、干旱事件）；以及 (4) 如何处理海量数据。

生态水文学 – 需要先进的人工智能架构，以开发新的数据产品和基准数据集，覆盖从微生物和叶片尺度到流域和大陆尺度的各种空间尺度。通过新颖的协同设计方法构建并收集基于过程的模型所需的带标签地球科学数据，并将其开源给 BER社区，将有助于快速测试现有的人工智能/机器学习方法。

气溶胶与云 ——能够提取有价值信息，或识别实际与模拟气候系统中强迫变化及涌现属性的指示性模式的协同设计方法至关重要。未来的系统概念若能构建用于存储指示性模式（例如冰成核或颗粒物、成雪）及涌现属性的数据库，将为知识发现开辟道路，并揭示必须纳入过程模型的缺失机制。

海岸动力学、海洋和冰 – 能够改进以下方面的先进人工智能架构：(1) 异构数据集的标准化与融合；(2) 过程模型中的尺度感知与依赖关系（例如，利用稀疏数据集，捕捉跨尺度的海岸、海洋和冰冻圈过程）。

气候变率与极端事件 – 气候变率、信号识别和可预测性来源的协同设计方法至关重要。这些方法包括用于检测与热带气旋、锋面、大气河流、冰雹、龙卷风和冰暴相对应的特征信号和模式的人工智能架构。

人类系统与动力学 – 能够加深对人类系统和地球系统理解的协同设计方法。例如，为了更深入地洞察极端事件背景下的城市预测和长期城市政策，需要发展人工智能架构。

6. 结论

在这篇观点论文中，我们阐述了采用协同设计方法，高效、准确地整合过程模型与观测数据，从而提升地球系统可预测性的必要性。当前的科学研究水平和高性能计算设施为应对“模型—实验”循环中的重大挑战提供了起点。我们需要构想未来的系统，将边缘传感器连接到智能计算设备，再连接到部署于雾计算／云计算／百亿亿次级计算基础设施中的过程模型，从而变革 ModEx 生命周期。我们的近期至长期目标使我们能够利用未来系统概念开发人工智能架构和协同设计方法。领域专家与计算专家携手开展社区协作，能够改变我们对综合性水循环及相关极端事件的建模方式。

术语表

- AI4ESP：地球系统可预测性人工智能
- 人工智能：人工智能
- 阿贡领导力计算设施：阿贡领导力计算设施
- ARM：大气辐射测量气候研究设施
- ASCR：先进科学计算研究
- ATS：先进陆地模拟器
- AWS：亚马逊云科技
- E3SM: 能源部百亿亿次级地球系统模型

• 地球系统模型：地球系统模型

• ESP：地球系统可预测性

• EMSL：环境分子科学实验室

• BER：生物与环境研究

• CPU：中央处理器

• 美国能源部：能源部

• GCP：谷歌云平台

• GPU：图形处理器

• 高性能计算：高性能计算

• 物联网：物联网

• LCF：领导计算设施

• ModEx：模型实验

• ML：机器学习

• 美国国家航空航天局：美国国家航空航天局

• 美国国家能源研究科学计算中心：美国国家能源研究科学计算中心

• 自然语言处理：自然语言处理

• 美国国家海洋和大气管理局：美国国家海洋和大气管理局

• 橡树岭领导力计算设施：橡树岭领导力计算设施

• 张量处理单元：张量处理单元

• 不确定性量化：不确定性量化

致谢。 作者感谢所有参与地球系统可预测性人工智能（AI4ESP）研讨会的人员所付出的努力。MKM 感谢环境分子科学实验室的支持。该实验室是美国能源部科学办公室用户设施，由生物与环境研究计划资助，合同编号为 DE-AC05-76RL01830。PJ 的研究是能源部百亿亿次级地球系统模型（E3SM）项目的一部分，获得美国能源部科学办公室和生物与环境研究办公室的资助。SS 和 MN 的研究获得百亿亿次计算项目（17-SC-20-SC）的支持。该项目由美国能源部科学办公室和国家核安全管理局合作开展。MBG 的研究由劳伦斯利弗莫尔国家实验室代表美国能源部开展，合同编号为 DE-AC52-07NA27344。MKM、JAA 和 MH 感谢太平洋西北国家实验室（PNNL）先进无线通信实验室的 Johnathan Cree 和 Elena Peterson，他们为本文制作了插图。本文由太平洋西北国家实验室（PNNL）撰写。该实验室由巴特尔纪念研究所运营，代表美国能源部开展工作，合同编号为 DE-AC05-76RL01830。本文由橡树岭国家实验室撰写。该实验室由 UT-Battelle有限责任公司运营，根据与美国能源部签订的合同开展工作，合同编号为 DE-AC05-00OR22725。美国政府保留非独占、已付清、不可撤销、全球范围内的许可权；出版商接受本文发表，即承认美国政府保留此项许可权，可出版本文的已发表版本或复制该版本，也可授权他人这样做，以满足美国政府的需要。美国能源部将根据美国能源部公共访问计划，向公众开放这些联邦资助研究成果。

数据可用性声明。本文未开发任何数据。

参考文献

Acer, S., 及合作者，2021：Exagraph：用于实现百亿亿次级应用的图与组合方法。《国际高性能计算应用杂志》，**35 (6)**，553–571。

阿贡领导力计算设施，2022：阿贡领导力计算设施 – 阿贡领导力计算设施。网址 https://www.alcf.anl.gov/，访问日期：2022-11-10。

ARM，2022：ARM – 大气辐射测量气候研究设施。网址https: //www.arm.gov/，访问日期：2022-11-10。

ATS，2022：ATS——高级陆地模拟器。网址：https://github.com/amanzi/ats，访问日期：2022-11-10。

AWS，2022：AWS——亚马逊云科技。网址：https://aws.amazon.com/about-aws/，访问日期：2022-11-10。

Azure，2022：Microsoft Azure：云计算服务。URL https://azure.microsoft.com/en-us/，访问日期：2022-11-10。

Baker, N. 和合著者，2019年：《科学机器学习基础研究需求研讨会报告：人工智能核心技术》。技术报告，美国能源部科学办公室（SC），美国华盛顿特区。https://doi.org/10.2172/1478744.

Barkjohn, K. K., B. Gantt, 和 A. L. Clements，2021：针对使用 PurpleAir 传感器采集的 PM 2.5 数据开发并应用一项全美范围的校正方法。大气测量技术，**14 (6)**，4617–4637。

Bauer,P., P. D.Dueben, T. Hoefler, T. Quintino, T. C.Schulthess, and N. P. Wedi, 2021: 地球系统科学的数字革命。自然计算科学，**1 (2)**，104–113。

Beckman, P., R. Sankaran, C. Catlett, N.Ferrier, R. Jacob, and M.Papka, 2016: Waggle：用于边缘计算的开放式传感器平台。*2016 IEEE* 传感器，IEEE，1–3。

Beckman, P. 和合著者，2020：5G赋能的能源创新：面向科学的先进无线网络（研讨会报告）。技术报告，阿贡国家实验室（ANL），美国伊利诺伊州阿贡；西北大学..... https://doi.org/10.2172/1606538。

Bringmann、O. 和其他作者，2021：面向边缘人工智能的自动化硬件/软件协同设计：现状、挑战与未来步骤：专题会议论文。*2021* 年国际硬件/软件协同设计与系统综合会议（*CODES+ ISSS*），IEEE，11–20。

Catlett, C., P. Beckman, N. Ferrier, H. Nusbaum, M. E. Papka, M. G. Berman, and R. Sankaran, 2020：利用软件定义的传感器测量城市。《社会计算期刊》，**1 (1)**，14–27。

Catlett, C. E.、P. H. Beckman、R. Sankaran 和 K. K. Galvin，2017：万物阵列：公共道路上的科学研究仪器：平台设计与早期经验。第二届智能城市运营与平台工程科学国际研讨会论文集，26–33。

Chambers, J.、R. Fisher、J. Hall、R. J. Norby、S. C. Wofsy 和 D. Stover，2012 年：《气候变化下热带生态系统研究重点研讨会报告》，2012 年 6 月 4—5 日。技术报告，美国能源部科学办公室（SC），美国华盛顿特区。生物学与……网址：https://ess.science.energy.gov/wp-content/uploads/2020/12/NGEE-Tropics3webHR.pdf。

Charles, Z.、Z. Garrett、Z. Huo、S. Shmulyian 和 V. Smith，2021：关于联邦学习的大规模队列训练。神经信息处理系统进展，**34**，20 461–20475。

Chen, K.M., E. M. Cofer,J. Zhou, andO.G. Troyanskaya,2019年：Selene：一个基于 PyTorch 的序列数据深度学习库。自然方法，**16 (4)**，315–318。

Chen, X., X. Huang, C. Jiao, M. G. Flanner, T. Raeker, and B. Palen, 2017：在商业云计算环境中运行气候模型：使用亚马逊 AWS 上的社区地球系统模型（CESM）进行的案例研究。计算机与地球科学，**98**，21–25。

Chouhan, L., P. Bansal, B. Lauhny, and Y. Chaudhary, 2020：云联邦架构及其挑战综述。社交网络与计算智能，Springer，51–65。

Christensen, B., and J. Blanco Chia, 2017：Raspberry Shake——全球公民地震仪网络。*AGU* 秋季会议摘要，第 2017 卷，S11A–0560。

Cortés, U., U. Moya, and M. Valero, 2021：当 Sally 遇见 Harry，或当人工智能遇见高性能计算。超级计算前沿与创新，**8 (1)**，4–7。

Cromwell, E.、P. Shuai、P. Jiang、E. T. Coon、S. L. Painter、J. D. Moulton、Y. Lin 和 X. Chen，2021：使用深度神经网络根据河流流量数据估算流域地下渗透率。地球科学前沿，**9**，613011。

Daepp, M. I. 和合著者，2022：Eclipse：面向城市低成本超本地环境传感的端到端平台。*2022*年第*21*届*ACM/IEEE*传感器网络信息处理国际会议（*IPSN*），IEEE，28–40。

Descour、M.，J.Tsao，D. Stracuzzi，A. Wakeland，D.Schultz，W. Smith和 J.Weeks，2021：工作坊报告：人工智能增强的下一代微电子协同设计：创新创新。技术报告，桑迪亚国家实验室（SNL-NM），美国新墨西哥州阿尔伯克基。https://doi.org/10.2172/1845383。

Di Lorenzo, P., C. Battiloro,M. Merluzzi和 S. Barbarossa，2021：无线网络边缘自适应联邦学习的动态资源优化。*ICASSP 2021-2021 IEEE* 国际声学、语音与信号处理会议（*ICASSP*），IEEE，4910–4914。

E3SM，2022：E3SM——能源部百亿亿次级地球系统模型。网址：https://e3sm.org/，访问日期：2022-11-10。

E3SM-MMF，2022：E3SM 多尺度建模框架。网址 https://www.exascaleproject.org/research-project/e3sm-mmf/，访问日期：2022-11-10。

EMSL，2022：《EMSL——环境分子科学实验室》。网址：https://www.emsl.pnnl.gov/，访问日期：2022-11-10。

GCP，2022：GCP——Google Cloud Platform。网址：https://cloud.google.com/，访问日期：2022-11-10。

Germann，T. C.、A. L. McPherson、J. F. Belak 和 D. F. Richards，2013：极端环境材料超大规模计算协同设计中心。https://doi.org/10.2172/1116965。

Halappanavar，M.、M. Minutoli 和 S. Ghosh，2021：超大规模计算时代的图分析。第*18*届 *ACM* 国际计算前沿会议论文集，209–209。

Heroux，M. A. 及合著者，2022：ECP 软件技术能力评估报告。技术报告，橡树岭国家实验室（ORNL），美国田纳西州橡树岭。https://doi.org/10.2172/1760096。

Hickmon, N. L., C. Varadharajan,F.M. Hoffman,S. Collis, andH. M. Wainwright,2022:地球系统可预测性人工智能（AI4ESP）研讨会报告。技术报告，阿贡国家实验室（ANL），阿贡，伊利诺伊州（美国）。https://doi.org/10.2172/1888810。

Hoffman，F. M. 和合著者，2017：2016 年国际陆地模型基准测试（ilamb）研讨会报告。技术报告，美国能源部科学办公室，美国华盛顿特区。https://doi.org/10.2172/1330803。

Lichtner, P. C. 和合著者，2020：PFLOTRAN 网页。Http://www.pflotran.org。Lie, S.，2021：数百万核心、多晶圆人工智能集群。*2021 IEEE*热芯片*33* 研讨会（*HCS*），IEEE 计算机学会，1–41。

Lindsey, N. J., T.C. Dawe, andJ.B.Ajo-Franklin,2019年：利用暗光纤分布式声学传感照亮海底断层与海洋动力学。科学，**366 (6469)**，1103–1107。

Morell, J. Á., 和 E. Alba，2022：利用志愿者边缘设备实现动态自适应容错异步联邦学习。未来一代计算机系统，**133**，53–67。

Mudunuru, M. K., K. Son, P. Jiang, G. Hammond, 和 X. Chen，2022：用于流域模型校准的可扩展深度学习。地球科学前沿，2206。

Mudunuru, M. K. 和合著者，2021：EdgeAI：如何利用人工智能收集可靠且相关的流域数据。技术报告，地球系统可预测性人工智能（AI4ESP . . . . https://doi.org/https://doi.org/10.2172/1769700。

美国国家能源研究科学计算中心，2022：美国国家能源研究科学计算中心。网址 https: // www.nersc.gov/，访问日期：2022-11-10。

橡树岭领导力计算设施，2022：橡树岭领导力计算设施 – 橡树岭领导力计算设施。网址 https://www.olcf.ornl.gov/，访问日期：2022-11-10。

Ott, J., M.Pritchard,N. Best, E. Linstead, M. Curcic, and P. Baldi,2020：用于科学计算的 Fortran-Keras 深度学习桥接工具。科学编程，**2020**。

Reuther, A., P. Michaleas, M. Jones, V. Gadepally, S. Samsi, 和 J. Kepner，2021：人工智能加速器调查与趋势。*2021 IEEE* 高性能极限计算会议（*HPEC*），IEEE，1–9。

Rosa, M. J.,C. G. Ralha,M.Holanda,和 A.P. Araujo，2021：面向联邦云中科学工作流的计算资源与成本预测服务。未来一代计算机系统，**125**，844–858。

Saxena、D.、R. Gupta 和 A. K. Singh，2021：《多云架构综述与比较研究：云联邦面临的新兴问题与挑战》。*arXiv* 预印本 *arXiv:2108.12831*。

Stavroulas, I. 和其他作者，2020：在希腊不同城市空气质量条件下对低成本 PM 传感器（PurpleAir PA-II）的现场评估。大气，**11 (9)**，926。

Stevens,R., V. Taylor, J. Nichols,A. B. Maccabe, K.Yelick, and D.Brown, 2020：人工智能助力科学：美国能源部（美国能源部）关于人工智能助力科学的城镇厅报告。技术报告，阿贡国家实验室（ANL），美国伊利诺伊州阿贡。https://doi.org/10.2172/1604756。

Subsurface-ECP，2022：Subsurface：一个用于耦合流动、传输、反应和力学的百亿亿次级地下模拟器。网址：https://www.exascaleproject.org/research-project/subsurface/，访问日期：2022-11-10。

Talsma, C., K. C. Solander, M. K. Mudunuru, B. Crawford 和 M. Powell，2022：利用机器学习和深度神经网络模型预测霜冻，应用于物联网传感器。

Testbed, A. A., 2022：阿贡领导力计算设施人工智能测试平台，新一代人工智能加速器计算机。网址：https://www.alcf.anl.gov/alcf-ai-testbed，访问日期：2022-11-10。

Tsai, W.-P., D. Feng, M. Pan, H. Beck, K. Lawson, Y. Yang, J. Liu 和 C. Shen，2021：从校准到参数学习：利用大数据在地球科学建模中的规模效应。自然通讯，**12**，1–13。

Vetter, J. S. 和合作者，2022：极端异质性：2018 年极端异质性时代的高效计算科学：美国能源部科学办公室高级科学计算研究项目极端异质性研讨会报告。https://doi.org/10.2172/1473756。

Winter,K., D. Lombardi,A. Diaz-Moreno, 和 R. Bainbridge，2021：使用RaspberryShake监测东南极洲的冰震。《地震学研究快报》，**92 (5)**，2736–2747。

Xu, H., W. Wei, J. Dennis, and K. Paul, 2019年：在地球系统模型中使用云友好型数据格式。*AGU*秋季会议摘要，第2019年卷，IN13C–0728。

Yang,C.和合作者，2016：用于非静力大气动力学的10m 核可扩展全隐式求解器。*SC' 16*：国际高性能计算、网络、存储与分析大会论文集，IEEE，57–68。

Yu、Y. 和合著者，2022：利用异构众核架构计算表征地球系统模拟的不确定性。《地球科学模型开发讨论》，1–23。

Yuan、B.及合作者，2019年：利用机器学习从嘈杂环境中辨识喷发：以奇马约、新墨西哥州的二氧化碳驱动的冷水间歇泉为例。《地震学研究快报》，**90 (2A)**，591–603。

Zhang, S.和合作者，2020：在异构众核超级计算平台上优化高分辨率社区地球系统模型。《地球科学模型开发》，13 (10)，4809–4829。

Zhang, X.,W.Jiang,Y. Shi,和J. Hu，2019年：当神经架构搜索遇上硬件实现：从硬件感知到协同设计。*2019*年 *IEEE*计算机学会年度超大规模集成电路研讨会（*ISVLSI*），IEEE，25–30。

Zhu, T.,J.Shen,andE. R.Martin,2021：通过电信光纤传感器感知地球与环境动力学：美国宾夕法尼亚州的一项城市实验。《固体地球》，**12 (1)**，219–235。