

× 人工智能系统在工作

不断变化的职业心理社会环境

作者 / 塔希娜·卡拉莫娃





署名权4.0国际 (CC BY 4.0)

本作品受Creative Commons Attribution 4.0 International许可。参见：<https://creativecommons.org/licenses/by/4.0/> 用户可以根据许可协议的详细规定，重复使用、分享（复制和再分发）、改编（混搭、改编和在此基础上进行创作）这些内容。用户必须明确注明国际劳工组织（ILO）是该材料的来源，并说明是否对原始内容进行了修改。禁止在翻译、改编或其他衍生作品中使用国际劳工组织的徽章、名称和标志。

归属权 用户必须指明是否进行了更改，并按照以下方式引用作品：Karimova, T. *AI Systems in Action: A Transforming Psychosocial Work Environment* 国际劳工组织工作论文170。日内瓦：国际劳工局，2026年。© 国际劳工组织。

翻译。

- 如果对本作品进行翻译，必须在署名信息之外，同时添加以下免责声明：本译文是国际劳工组织（ILO）受版权保护的作品的翻译版本。本翻译未经国际劳工组织准备、审查或批准，不应被视为国际劳工组织的官方翻译。国际劳工组织对本译文的内容和准确性不承担任何责任。本译文的责任完全由译者承担。

适应 若对此作品进行改编，必须添加以下免责声明，并附带署名：这是联合国国际劳工组织（ILO）所创作的版权作品的改编版。这次编著未经联合国国际劳工组织编写、审阅或推荐，并不属于联合国国际劳工组织正式编写的材料。联合国国际劳工组织宣布免对本内容和准确性承担任何责任，责任全归编写改编的作家。

第三方材料 这个Creative Commons许可不适用于本出版物中包含的非国际劳工组织版权材料。如果材料归属第三方，使用此类材料的用户对与版权所有者的协商权利以及任何侵权索赔负全部责任。

任何在此许可下产生的、无法友好解决的争议，应按照联合国国际贸易法委员会（UNCITRAL）的仲裁规则提交仲裁。各方应受仲裁裁决的约束，该裁决是对此类争议的最终裁决。

有关版权和许可详情，请联系：rights@ilo.org
rightsi@ilo.org 有关国际劳工组织出版物和数字产品的详细信息，请访问：www.ilo.org/publns
国际劳工组织出版物。

ISBN 9789220434079（印刷版），ISBN 9789220434086（网络PDF版），ISBN 9789220434109（epub版），ISBN 9789220434093（html版）。ISSN 2708-3438（印刷版），ISSN 2708-3446（数字版）。

<https://doi.org/10.54394/00034098>

所使用的名称在ILO出版物中与联合国惯例一致，对其中材料的表述并不代表ILO对任何国家、地区或领土的法律地位的任何意见。

或者关于其权力机构，或者关于其边疆或边界的界定。参见：www.ilo.org/免责声明。

本出版物中表达的观点和观点是作者的观点，并不必然反映国际劳工组织的观点、观点或政策。

对公司和商业产品及工艺名称的提及并不代表国际劳工组织对其的认可，未提及特定公司、商业产品或工艺也不代表否定。

信息关于国际劳工组织的出版物和数字产品可在以下网址找到：www.ilo.org/研究与出版物

国际劳工局工作论文总结了国际劳工局正在进行的研究成果，旨在激发对与工作世界相关的一系列问题的讨论。欢迎对这篇国际劳工局工作论文提出评论，评论可以通过以下方式发送：research@ilo.org
[研发@ilo.org](mailto:research@ilo.org)。

授权发布：卡罗琳·弗雷德里克森，研究部总监

ILO工作论文可在以下网址找到：www.ilo.org/research-and-publications/working-papers

建议引用：

卡拉莫娃，T. 2026. *AI Systems in Action: A Transforming Psychosocial Work Environment* 国际劳工组织工作文件 170 (日内瓦，国际劳工组织) <https://doi.org/10.54394/00034098>

摘要

工作场所中人工智能系统的主导框架集中在基于人工智能的技术如何改善工作场所的安全和健康方面。很少有例外，对这些技术的负面影响很少进行映射和理解。尽管如此，对需要批判性地审查现有职业安全与健康框架提供的预防方法是否适合应对与在职场中部署基于人工智能的系统相关的风险的意识正在逐渐增强。关于这个话题的辩论在监管层面非常活跃。各个司法管辖区正在制定一般人工智能法规，这些法规倾向于将就业环境中基于人工智能系统的使用归类为高风险。本文旨在考察人工智能对健康的影响，重点关注其对心理健康和社会福祉的影响（在职业安全与健康学科中被称为工作场所心理社会环境）。本文认为，为了有效地应对由基于人工智能的系统产生的心理社会风险，政策制定者应采取包括劳动和就业法、平等和非歧视法、职业安全与健康法以及隐私和数据保护法在内的综合性方法。

关于作者

塔希玛·卡拉莫娃 是一位专门从事国际公法、可持续发展、国际劳工标准、人权法和武器控制的律师。她是国际劳工组织研究部门的法律研究专家。此前，她在人权事务高级专员公署担任人权官员，在此之前，她在日内瓦国际人道法与日内瓦大学法学院人权/法学系担任研究助理。她持有埃塞克斯大学的法学硕士和日内瓦大学及日内瓦国际与发展研究院的国际法博士学位。

目录表

摘要 01
关于作者 01

X	序言	03
---	----	----

X	1 布置场景	05
	人工智能系统和它们在工作中的用途	05
	健康与安全在基于AI的工作场所	06

X	2 心理社会风险与人工智能系统：相互作用和数据可获得性	09
	工作场所的心理社会因素	09
	人工智能技术引发的心理社会风险：研究现状	10
	关注新的或“增强型”风险	12
	密集/侵入式监控	13
	工作自主性（及尊严）	14
	过度收集数据和缺乏透明度	16

X	3 监管框架针对人工智能相关PSR	17
	通用人工智能法规及其中的职业健康安全范围	19
	国家针对基于人工智能的风险的职业安全健康相关法规应对措施	24
	“新”职业健康安全权利的出现	28

X	结论	30
---	----	----

	参考文献	31
--	------	----

X 序言

工作领域和工作性质正因数字化，包括基于人工智能的技术，经历着重重大变革。¹ 在全个工作场所，人工智能（AI）正越来越多地被用作工具，用于完成工人的任务，但也涉及到雇佣关系的整个周期，包括招聘、培训、入职、监控和监督、薪酬、奖励以及职业发展、调动和解雇。这种变革正在深刻改变劳动力的构成、工作环境、使用的设备，以及工作的设计和组织方式，对工人的健康和福祉产生重要影响。²

由于这些变革，学者们开始分析数字化和新型人工智能技术的影响。³ 政策制定者和从业者努力更好地理解如何在工作中治理技术；而区域性机构已经开始发起活动并研究数字化的工作世界以及这对工人的意义。关于人工智能系统如何融入雇佣关系核心方面（如工作时间、工资、职业安全与健康（OSH）以及其他工作权利）的研究仍然有限。⁴

为了解决这一差距，我们需要探讨各种问题，包括工作的意义，以及旨在确保“所有人都享有人道的工作条件”的政策是否适合应对这个以数字和人工智能为基础的工作世界的挑战。目前，关于工作场所人工智能技术的讨论主要集中在这些技术带来的机遇上，而且除少数例外，关于人工智能技术对健康影响的知识体系化程度很低。因此，本文将不涉及工作场所人工智能应用的积极进展或从职业健康安全（OSH）角度分析其益处。相反，它将关注日益增长的关于人工智能系统对健康和福祉影响的研究证据和担忧。⁵ 包括人们不太理解的职场安全与健康方面，例如心理社会风险——定义为“在工作设计或管理中增加职业相关压力风险的一切因素”。⁶ 这种担忧既体现在更广泛的政策制定背景下，与人工智能技术的整体监管相关，也体现在与新的就业特定标准相关的领域，例如禁止某些基于人工智能的职场监控形式，以及在人工智能算法管理（AM）背景下声称的“新”或“改进的旧”权利（例如，对人工智能技术部署的透明度权利，对基于人工智能的AM所作决定的 humans review 权利等）。

这篇研究论文探讨了是否需要将现有的法律框架调整以适应与工作场所中人工智能部署相关的新的健康风险，以及这些调整应该是什么。因此，它探讨了数字化工作环境如何激发心理社会问题。

¹ 根据国际劳工组织（ILO）的说法，虽然“数字化涵盖工作场所中所有类型的数字应用，从简单的桌面计算机使用到高级机器人和虚拟/增强现实，自2010年代中期以来，关注的焦点主要集中在人工智能，尤其是机器学习”。《数字化面临的挑战与机遇》，GB.350/HL/1(修订1)，§5。

² C. de Tecco, B. Persechino, S. Iavicoli, 《在变化的工作世界中关注心理社会风险：从风险评估文化转向机遇管理》《Med Lav. 稍微劳，或者 中度疲劳》，2023, 114(2)。

³ A. Aloisi & V. De Stefano, 《你的老板是算法：人工智能、平台工作和劳动》（布卢姆斯伯里出版社，2022年）。

⁴ 参见例如欧盟-欧莎健康工作场所活动：《安全健康地工作在数字时代》，2023-2025。

⁵ M. Jarota, 《工作过程中的人工智能》。从职业健康安全视角对欧盟关于人工智能拟议法规的反思《计算机法律与安全评论》卷49，2023年。

⁶ 国际劳工组织《心理社会风险和工作压力》，2022年11月10日。

危害以及这些危害是否需要额外的和/或不同的安全措施来保护工人的健康和安全。⁷ 通过这样做，该论文将有助于填补已知和新兴风险因素的知识空白，以确保工作场所的安全、健康和福祉，并帮助企业更好地理解与人工智能技术相关的具体风险，以推动体面劳动的发展。

该研究首先确立了人工智能技术在职场和 workflows 中的背景，并介绍了本文所探讨的AI技术类型（第1节）。在此基础上，分析了这些技术对工作安全与健康的影响。第2节探讨了与这些AI技术相关的新的和潜在的健康和安全风险，重点关注由基于AI技术的心理社会风险。然后，文章在第3节中考察了国际劳工标准以及区域和国家劳工监管趋势的作用。文章主张需要一个综合的监管方法，将预防方法与新兴的监管干预措施相结合，以解决与基于AI系统的心理社会风险相关的问题。本文聚焦于职场安全与健康的一个具体方面——职业安全风险（PSRs），并将不涉及身体健康和安全性问题。

⁷ F. Chirico，被遗忘的新兴精神社会风险因素的领域。职业健康

× 1 布置场景

人工智能系统和它们在工作中的用途

在探讨基于人工智能的技术与工人健康和安全性之间的实际互动之前，重要的是要定义什么是人工智能，以及它在工作场所和工作流程中的使用方式。

有各种方式来定义“人工智能”这一术语。简单来说，这一概念被定义为“机器思考、学习和适应的能力”，并且重要的是，人工智能“不再局限于常规任务，现在它还应对着曾经只有人类智慧才能解决的复杂挑战”。⁸ 监管框架同样试图定义人工智能。例如，最近通过的欧盟人工智能法案将人工智能系统定义为“一种基于机器的系统，旨在以不同的自主程度运行，并在部署后可能表现出适应性，并且，为了明确或隐含的目标，从其接收的输入中推断出如何生成预测、内容、推荐或决策等输出，这些输出可以影响物理或虚拟环境”。⁹

基于人工智能的技术——一个需要其自身定义的术语——是更广泛的工作场所数字化进程的一个子集，这个过程涵盖了广泛的技术，如人工智能、高级机器人技术、远程工作监控技术、物联网（IoT）、大数据、可穿戴设备和在线平台等。¹⁰ 本文聚焦以下类型的技术：高级机器人、基于人工智能的增材制造（见框）。

非常简短的先进机器人、基于人工智能的增材制造和智能数字系统简介

高级机器人（包括AI赋能系统）¹¹ 或智能机器人是“专为执行需要高精度、适应性和自主性的任务而设计的技术……[例如]工业机器人、……用于重复性和危险任务的机械臂，以及自主移动机器人、无人机、外骨骼和协作机器人（“协作机器人”）等现代创新”。¹² 这些基于AI的机器人在先进制造中的应用越来越广泛。

基于AI的AM（也称为“算法化管理”）可以被定义为在组织中利用算法程序协调劳动力投入的方法。¹³ 但是，它也可以被定义为将管理职能委托给算法。¹⁴ 基于AI的AM从工作场所、工人及其活动中收集数据。这些数据随后由基于AI的系统进行处理，“以做出自动化或半自动化决策，或为决策者提供信息，例如人力资源经理、雇主，有时甚至是工人本人”。¹⁵

⁸ ISO 什么是人工智能（AI）？参见牛津高阶英汉双解词典，其中将AI定义为“计算机或其他机器展现或模拟智能行为的能力”，见OED。人工智能。

⁹ 欧盟、人工智能法案，2024年3月13日，第3条。

¹⁰ 欧盟职业安全与健康局 数字技术在工作中的应用与心理社会风险：对职业安全与健康的证据与影响，2024。

¹¹ J. T. Licardo, M. Domjan, T. Orehovački, 智能机器人——对新兴技术和趋势的系统回顾 电子学 13(3)。

¹² 国际劳工组织 革命化健康与安全：AI和数字化在工作中的作用 2025，第6页。

¹³ S. Baiocco, E. Fernandez-Macias, U. Rani 和 A. Pesole 工作算法管理及其在不同环境中的影响，（国际劳工组织，2022年），第5页。

¹⁴ 麻省理工学院斯隆管理评论 算法管理：人工智能在管理团队中的作用，2023年4月。

¹⁵ 欧盟职业安全与健康局 数字技术在职场中的应用与心理社会风险 第24页。

智能数字系统也可以整合基于AI的技术。在工作场所环境中，这些系统包括“使用一系列数字技术——如基于传感器的设备、AI、物联网、可穿戴设备、无线技术、增强现实、虚拟现实和无人机——来监测、分析和管理工作场所的安全和健康风险，包括与工人活动或任务、设备、工作场所布局和工作组织等因素相关的物理、人体工程学、化学、生物和心理社会风险”。¹⁶

尽管这些技术在许多方面存在重叠，但以AI为基础的技术位于每个技术的核心。因此，文献根据它们如何改变工作世界来对它们进行分类。讨论区分了AI技术在职场中的两种主要应用：

□□ 第一个目的是自动化工人执行的任务，特别是那些可以通过机器高效处理的日常和重复性任务。第二个目的是利用基于AI的分析和算法来取代或增强管理职能：招聘、监控、监督和培训员工，以及安排工作时间和休息时间——通常被称为“算法管理”。两者都对就业数量（就业岗位数量）和就业质量有影响，包括尊重工作中的基本原则和权利。¹⁷

现有文献进一步丰富了人工智能在工作场所的应用范围，通过记录人工智能技术在工作场所的其他应用，例如在评估员工绩效中使用人工智能技术。¹⁸ 监控和预测员工健康状况¹⁹ 甚至直接改善员工健康。²⁰ 这些例子说明了工人与人工智能之间之前未知的关系的存在。

这项研究将“AI系统”和“AI技术”互换使用，以描述考察的三种技术类型（即高级机器人、基于AI的增材制造和智能数字系统）。

健康与安全在基于AI的工作场所

研究人员发现，基于人工智能技术的数字化工作场所为工人（例如，自动化常规和重复性任务，降低身体风险）和雇主（例如，提高生产力和效率）以及改善工作场所安全提供了多种机会。²¹ 然而，这些机会同时也伴随着对工人安全与健康以及福祉的各种风险，包括生理、心理、社会心理、经济和道德风险。²² 说到这里，正如所述，这项研究将仅关注PSRs。随着基于人工智能的技术在各行各业的高度集成以及涵盖整个雇主/管理职责范围，它们对工作场所健康的一个直接影响将在心理社会因素层面感受到。其次，PSRs的研究较少，因此相对而言了解得较少。

合作机器人在工人旁工作时或与工人分享一项任务时，可能会给工人带来创伤性损伤或认知超载以及工作强度增大的风险。

¹⁶ 欧盟职业安全与健康局 [数字技术在职场中的应用与心理社会风险](#)，第19页。

¹⁷ 国际劳工组织 [数字化面临的挑战与机遇](#)，§7。

¹⁸ A. Varma, V. Pereira & P. Patel, 人工智能与绩效管理 *组织动力学* 第53卷：第1期（2024年）。

¹⁹ El-Helaly M. 人工智能与职业健康安全，利弊分析 *Med Lav. 稍微劳，或者 中度疲劳*。115(2)(2024)。

²⁰ 国际劳工组织 [革命化健康与安全：AI和数字化在工作中的作用](#)，2025，第9页。

²¹ For a broad overview see ILO。

对于概述，请参阅国际劳工组织（ILO）。 [革命化健康与安全：AI和数字化在工作中的作用](#)。

²² J. Howard & P. Schulte [管理职场人工智能风险与未来工作](#)，*美国工业医学杂志* 第67卷第11期（2024年），第4页。

在附近²³ 工作强度提升，例如，可能是由于“人类工人的体力或认知能力与协作机器人人工智能驱动的节奏之间不匹配”所导致。²⁴ 首先，对失业和职业不稳定的普遍恐惧一直是与先进数字技术“合作”的工人关注的焦点。²⁵

人工智能分析和算法的使用通常导致工人与人工智能系统之间的持续互动，影响工作的各个方面，尤其是日常任务的监督和管理。²⁶ 这些系统不再是实验性的，而是正在被整合到传统产业中。²⁷ 一项对六个高收入国家中层经理的调查显示，在6,047名被调查的经理中，关于其公司使用AM工具的情况，74%的人表示，他们的公司至少使用一种工具来指导、监控或评估员工。²⁸ 基于AI的AM假定管理权限²⁹ 大幅度改变传统的雇主和雇员关系，同时也可能带来对安全和健康工作产生不利影响的潜在因素。

新出现的证据表明，基于AI的自动化监控能力与工人遭受各种不良的心理社会和健康影响相关联。这可能会对工人的健康状况、工作质量和雇主与雇员关系产生“削弱和不等的”影响。³⁰ 该调查还发现，在六个高收入国家的管理层中，27%的中层管理人员担心员工的身心健康往往得不到充分保护。³¹

证据表明，自动化管理工具对工人在工作满意度、信任、工作量、动力和压力水平等方面产生负面影响。因此，另一项在美国进行的调查，使用1273名受访者（工人）的数据，发现46%表示“一直”被监控生产力（即使用自动化管理和监控技术）的工人认为他们工作速度过快，而只有15%表示从未被电子监控过生产力的工人这样认为。³² 根据相同的调查，表示他们的生产力“一直都在”被监控的53%的员工表示，他们时常或总是感到工作焦虑。这种焦虑可能与收集并输入人工智能的数据没有考虑上下文信息或关联到允许对员工生产力习惯做出合理结论的数据点有关。如澳大利亚议会就业、教育和培训常任委员会提交的文件中所述，像按键监控这样的过度监控，作为对员工生产力的“代理”，并不能讲述“完整的故事”。³³

²³ K. Jung & J.-S. Yang, 《缓解人机协作中的安全挑战：人类能力的作用》，《科技预测与社会变迁》，卷213（2025）。

²⁴ J. Howard & P. Schulte, 《管理职场人工智能风险与未来工作》，第4页。

²⁵ L. Tomidei 等人, 《超越纯技术——协作机器人对认知和组织的影响》，2022年12月12日；欧盟职业安全与健康局, 《数字技术在职场中的应用与心理社会风险》第15页。

²⁶ J. Howard & P. Schulte, 《管理职场人工智能风险与未来工作》，第4页。

²⁷ U. Rani, A. Pesole, I. Gonzalez Vazquez, 《算法化管理实践在常规职场中的应用：物流和医疗保健领域的案例研究》（国际劳工组织，2024）。Cox T. 和 Oosterwijk G. R., 《算法在职场中的应用：北欧七行业算法技术影响的案例研究》政策研究，2024年10月。

²⁸ 经合组织, 《职场算法管理：来自经合组织雇主调查的新证据》，《人工智能论文》，31（2025）。在国家层面：在美国，90%的经理表示他们的公司至少提供一种算法管理工具，而欧洲国家（法国、德国、意大利和西班牙）的流行率在76%（意大利）到81%（法国）之间”，第19页。

²⁹ GuruLink, 《算法管理：人工智能中层管理的崛起》。

³⁰ J. Howard & P. Schulte, 《管理职场人工智能风险与未来工作》，第5页。

³¹ 经合组织, 《职场算法管理：来自经合组织雇主调查的新证据》。

³² 赫尔特-费尔南德斯, 《评估工作中自动化管理和监控技术的普及及其对员工福祉的影响》，2024。

³³ 澳大利亚，国会就业、教育和培训委员会, 《工作未来》，2025年，第46页。

上述证据表明，人工智能是导致工作满意度、信任、工作量和动力等层面挑战的责任所在。就业关系的数字化发生在这样一个背景下：工作场所的心理社会风险（PSR）常常被忽视，全球范围内的大量立法都集中在工作场所安全物理方面，而非心理和社会健康方面。

X 2 心理社会风险与人工智能系统：相互作用和数据可获得性

工作场所的心理社会因素

一份关于国际劳工组织/世界卫生组织职业卫生联合委员会的报告将工作中的心理社会因素定义为“工作环境、工作内容、组织条件与工人能力、需求、文化以及个人工作外考虑因素之间的相互作用，这些因素可能通过感知和经验影响健康、工作表现和职业满意度”。³⁴

这个定义作为参考点，用于概括一般工作环境/条件与主观要素（如个人心理和健康能力）之间的动态相互关系。任何在工作和管理工作设计或管理中增加工作相关压力风险的因素，都可以被视为心理社会危害。³⁵ 换句话说，可能导致压力的工作场所因素被定义为心理社会危害。³⁶ PSRs和危害这两个术语通常被交替使用。

心理社会因素，也称为那些可导致压力的工作场所因素，因此变为工作中的心理社会隐患，包括：³⁷

- 工作内容/任务设计（工作缺乏多样性；技能使用不足或缺乏适合工作的技能）；

工作负荷和工作节奏（长时间或不社交的工作时间；轮班工作；不灵活的工作时间）

- 职业控制（对职业设计或工作负荷缺乏控制；有限参与决定自己的工作方式自主作品）；

环境与设备（不安全设备与资源；恶劣的物理工作条件，如照明不足、噪音过大或刺耳、不合理的工效学设计）；

- 组织文化（不明确的组织目标；沟通不畅；允许歧视或虐待的文化）

工作中的人际关系（社会或物理隔离；来自上司或同事的有限支持；权威性的监督和不良的直线管理；暴力、骚扰或欺凌；歧视和排斥）

- 在组织中的角色（组织或团队中不明确的职位角色）
- 职业生涯发展（晋升过快或过慢；工作不稳定；对发展的投入不足；对病假和绩效管理采取惩罚性措施）；及
- 家庭与工作界面（家庭与工作要求冲突；因工作远离家乡）。

³⁴ 国际劳工组织 工作场所的心理社会因素：识别和控制 《1986年国际劳工组织/世界卫生组织职业健康联合委员会报告》

³⁵ 国际劳工组织 通过职业安全与健康措施预防和处理工作中的暴力和骚扰，2024年，第12页。

³⁶ V. Forestieri，预防和减少心理社会风险及工作相关压力 《国际劳工研究杂志》，8(1-2)，1-131，第13页。

³⁷ 国际劳工组织 心理健康与社会风险以及工作压力，2022。

心理社会风险影响工人的福祉和生产力，可能导致如抑郁症、焦虑症、心血管疾病和肌肉骨骼疾病等身体和心理健康问题。此外，研究表明职业伤害与心理社会风险之间存在动态关系。³⁸

在多个国际劳工标准中，有关精神健康的条款中都提到了PSR。首先，1981年国际劳工组织职业安全与健康公约（第155号，以下简称C155）将“健康”定义为与工作相关，不仅指没有疾病或虚弱，还包括直接影响工作安全和卫生的生理和心理因素。³⁹此外，《C155》第五条规定，应将机械设备、工作时间、工作组织和工作流程适应工人的身心能力（着重强调），强调“良好的心理健康、良好的工作条件、足够的薪酬、工作效率、幸福感和生活质量”之间的正相关关系。⁴⁰

职业安全与健康建议，1981年（第164号），在C155的第4条中亦提到，因工作条件导致的有害身心压力需要预防。尽管这些措施属于政策层面，但在企业层面，雇主有责任不仅在一般意义上，“提供和维护工作场所、机械和设备，并采用工作方法，以确保”安全且对健康无害“合理可行”但也要“采取一切合理可行的措施，以消除过度身心疲劳”。⁴¹

国际劳工组织关于工作中暴力与骚扰的第190号公约（以下简称C190）及其伴随的建议书明确处理了PSRs，后者通过暴力和骚扰的概念来阐述。虽然第1条将心理伤害纳入工作场所暴力和骚扰的定义中，但C190的第9条设想了预防PSRs。鉴于第3条明确规定公约适用于在工作过程中、与工作相关或因工作产生的暴力和骚扰，包括通过工作相关通讯以及由信息和通信技术支持的那些，可以认为C190也适用于工作场所使用基于AI技术的风险。

人工智能技术引发的心理社会风险：研究现状

结合上述关于基于AI技术的互动与工作场所安全和健康的描述，以及当前的心理社会因素分类（也称为影响工作场所心理社会环境的工作特征），可能有助于更好地理解AI如何影响这些因素。（表1）

³⁸ J. Mireia, C. Catalina-Romero, E. Calvo-Bonacho & G. Benavides, 工作场所心理社会风险因素暴露与职业伤害发生率：西班牙一项队列研究。《职业与环境医学》，2016，(3)，第282-286页。另见S. Leka, A. Jain, S. Iavicoli & C. Di Tecco, 《欧盟工作场所心理社会风险和心理健康政策背景评估：成果、挑战与未来》。《生物医学研究国际》(2015)。

³⁹ 职业安全与健康公约，1981年（第155号），第3条；职业安全与健康建议，1981年（第164号），第2款。

⁴⁰ 欧洲议会2022年7月5日关于数字职场心理健康决议（2021/2098(INI)），第11段。

⁴¹ 职业安全与健康建议，1981年（第164号），第3段和第10段。心理社会因素也在1985年（第161号）《职业卫生服务公约》的框架中得到关注，该公约在第一条(a)(ii)款中定义了职业卫生服务，即向雇主、工人及其代表提供关于“建立和维护一个安全健康的工作环境，以促进工作相关的最佳身心健康的”要求，以及“根据其身心状况，适应工人的工作能力”。

X 表1：心理社会风险与数字技术特定风险分类（与人工智能相关的心理社会风险因素
载体）

心理社会因素或 工作层面的影响因素 心理学社会环境 工作态度	描述	AI系统带来的风险
职位内容/任务设计	工作缺乏多样性；技能利用不足 或者缺乏适当的工作技能。	Deskilling or the need for upskilling/re- 技能提升
工作量和节奏	长时间或不社交的工作时间；轮班制；在室内 弹性工作时间	认知超负荷，持续的必需 适应新技术；时间/ 节奏压力。 工作强度增加；自动化管理 引导致数量不断增加 任务在更短的工作时间内完成；无法实现 能或不合理的核心关键表现 指标。
工作控制 子类别：au损失 音域，意义 工作中的	缺乏对工作设计或工作量的控制 有限的参与决定自己的 工作	工人自主性降低；AM缺乏 透明度 能力做出决定，失去 有意义的职业，价值感丧失 感受降低的自我效能感，更强烈的 工作异化
环境和装备- 精神 子类别：AI 故障 问题；伤害	不安全的设备和资源；差的身体素质 ical working conditions, such as poor light- 优良的工 噪声，过度或令人烦恼的，差的功效学 经济学	员工中存在的公平感缺失 低条做集例对数据黑暗的 工作场所、工人及其活动。 工人可能因缺乏... 整理来自人工智能系统的信息，“算法” 节拍选择存在严重问题 由于部分或不完整的数据，不充分 quotemodelling, and problematic objective tive”.
组织文化 子类别：侵入性或 过度监视	不明确的企业目标，差劲的沟通 沟通；赋予歧视的文明 侵害或滥用。	由于数字监控导致的不信任；前 持续监控；过度收集 数据关于工人；隐私泄露，未经..... 授权收集和使用个人数据 数据收集的有效性；错误内部的风险 关于工人及其表现的不同意见 他们的个人怪癖之外 他们职能的范围 算法与情绪监测的生物监控 ment：心态；intention：意图；activity：活动；attention：注意力；select 工人行为的剥削

心理社会因素或工作层面的影响因素 心理学社会环境 工作态度	描述	AI系统带来的风险
人际关系 船只在工作	社交或身体隔离；支持有限 来自上级或同事的；权威性的 监督不力和一线管理薄弱；vio- 暴力，骚扰或欺凌；歧视 以及排斥。	人际关系不佳； 面对面交流的减少 工作场所凝聚力的削弱；增加 社交隔离
组织角色	组织内部不明确的职位角色 队伍	角色模糊：工人的身份，即如何 一个人在他们的职业中看待自己 职业角色
职业生涯发展	晋升不足或过度；工作不稳定 劣质发展投资；惩罚性的 sick leave procedures and perfor- 管理舞会	影响工人的经济流动性 工资下行压力；必要 转行；失业。
家庭-工作界面	冲突的家庭作业要求；离家在外 从家出去工作。	有限的断开能力；角色暧昧 愧疚

来源：国际劳工组织1984年、1995年，欧盟-欧洲职业安全与健康局2024年，J. Howard，P. Schulte 2024年，以及作者分析。

表1借鉴了ILO/WHO报告对心理社会因素的分类，并试图证明心理社会因素——尽管在原则上与技术无关——可能会因技术的设计和使用而引发PSR。一些由基于AI技术的使用引发的PSR，如过度监控和广泛的数据化，可以归类为属于现有因素，如组织文化、环境或设备的风险。然而，与此同时，这些风险也可以被单独识别为属于独特的、由新技术赋能的心理社会风险（例如，过度监控或监视）。

同样，在一个决策（例如，招聘、晋升、解雇、纪律处分）由算法做出的组织文化中，对工作的缺乏控制会引发工作场所自主性和潜在在工作场所歧视的问题。在数字化的背景下，“工作控制”这一概念获得了全新的意义。当前对工作心理社会方面的框架可能还没有充分考虑到与基于人工智能的技术和AM相关的PSR。可能需要制定新的心理社会因素/方面，以捕捉与技术相关的风险，例如员工对个人数据使用的控制。下一节将更详细地讨论由此产生的AI技术带来的“新”工作心理社会方面的一些内容。

关注新的或“增强型”风险

存在一些由AI系统部署产生的风险因素，它们并未精确地归属上述提到的心理社会因素类别之一，尽管它们似乎与（有时重叠）其中一些因素相关。这些显著的新PSR包括过度监控和监视、数字化工作场所中的工作自主性以及过度数据收集和缺乏透明度。下文旨在详细阐述这些“新”风险，并更好地理解其职业安全与健康影响。

密集/侵入式监控

在工作场所，对绩效的监控是标准做法。过去十年的技术发展，以及“管理文化”和“新的组织形式”，都扩大并加剧了工作场所“监视”的机会。⁴² 在这个背景下，心理社会风险因素可以“与监控的设计和使用相关，以及与围绕其的管理流程和政策相关”。⁴³ 在当代技术背景下，从语义上讲，“监控”一词指的是收集信息（“监控行为总是涉及有目的地收集关于某事或某人的信息”）。⁴⁴ 但是最近的政策使用了“过度/侵扰性监控/监视”这一术语。⁴⁵

并非所有数据收集或监控/监视都被视为未经授权。事实上，在某些情况下，职场监控通常是被授权的。正如所指出的，关键问题是强度、侵入性和不断通过技术实现的监控，这引发了职业安全与健康问题以及其他法律关切。随着数字技术的发展，关于员工及其工作所收集的数据量庞大且持续增加。例如，广泛使用的职场应用（如Microsoft Teams）可以生成有关用户活动数据，例如用户何时在电脑上活跃、在共享频道中的活动、在团队聊天或私人聊天中发布的消息、他们参与的电话、用户组织的会议数量、用户参与的会议数量、会议的音频、视频和屏幕共享时间、用户的最后活动日期、用户的共享频道互动等。⁴⁶

从职业健康安全角度来看，“根据配置方式的不同，监控也可能导致消极的隐私、信任、程序和分配正义感知，以及自主性和创造力的负面影响”。⁴⁷ 在一则报道中，“当技术无法被推翻，那些受到监控的人被剥夺了发表评论/纠正/陈述自己立场权利时，工作场所的尊严受到损害”。⁴⁸ 此外，“过度监控和监督以追踪员工实现关键绩效指标（KPI）的成果，可能会导致不安全的工作环境”，因为由人工智能系统产生的监控材料可能会导致关于员工表现的错误推断。⁴⁹

在欧盟层面，欧洲议会于2022年7月5日通过的关于数字工作世界心理健康的决议指出，“使用数字工具，如应用、软件和人工智能（AI）管理员工”，对工人的健康和安全，尤其是他们的心理健康、隐私权和人权尊严产生了“显著的危害风险”。AM为未来的工作提出了新的挑战，例如通过预测和标志工具实现的科技辅助控制和监视、远程实时监控进度和绩效以及时间追踪。决议指出，“现在约40%的国际公司人力资源部门使用人工智能应用，70%将其视为组织高度优先事项”。⁵⁰

在工作场所与技术相关的心理社会风险被国家政策认可的例子极为有限。一个典型的例子是澳大利亚的Comcare（国家健康与安全管理局）。

⁴² K. Ball, 《电子监控与工作场所监视》, (欧盟, 2021), doi:10.2760/5137, JRC125716, 第7页。

⁴³ 球, 第7页。

⁴⁴ 球, 第10页。

⁴⁵ 澳大利亚Comcare: 侵扰性监控。

⁴⁶ 参见 <https://learn.microsoft.com/en-us/microsoftteams/teams-analytics-and-reports/teams-reporting-reference>

⁴⁷ 球, 第7页。

⁴⁸ 澳大利亚《工作未来》, 第53页。

⁴⁹ 澳大利亚《工作未来》, 第52页。

⁵⁰ 欧洲议会《关于数字工作领域中心理健康的2022年7月5日决议》, (2021/2098(INI)), 第F段。

工作健康与安全），该机构已制定了一份符合侵入性监控标准的工作场所实践清单：

- 不合理的高监督/监管水平；
- 跟踪工人受到的方法和时间，如键盘活动追踪器；
- 监控电子邮件、文件和互联网使用情况；
- 通过工作电脑上的摄像头进行的秘密监视；
- 通过闭路电视和可追踪设备跟踪工人的通话和移动；
- 允许PCBU远程访问（未经工人知情/许可）并截取工人电脑屏幕的技术；
- 用以监控工作表现而非其他原因（如安全考虑），对公司车辆中工人的GPS监控工人移动情况。⁵¹

重要的是，在2024年澳大利亚工作健康与安全（工作场所心理社会危害管理）实践指南中，侵扰性监控被认定为工作场所危害。根据该实践指南，侵扰性监控方法/工具“可以是模拟的（例如，管理者进行微观管理）或技术的（例如，可追踪设备）”。⁵² 根据法规，侵入性监控可能包括不合理的监督水平；过度监控工作任务或休息时间；为绩效监控目的跟踪工人的通话和行动（使用监控摄像头或可追踪设备），而不是为了提高安全或其他目的；对工人使用公司车辆时的地理位置或移动进行GPS监控，目的也是为了绩效监控，而不是为了提高安全或其他目的。这些广泛的类别包括基于人工智能的广泛监控和监督。

尽管例子有限，监管环境可能正在缓慢变化，各国司法管辖区正在制定法律提案，如第3节所述。

工作自主性（及尊严）

工作控制是指工人对其工作的自主性和影响力的程度，它被认为是良好工作条件和职场福祉最重要的、如果不是根本的方面之一。对工作设计或工作量几乎或完全没有控制，以及/或者在对自己的工作做出决策中参与度有限，都是已知的工作场所心理社会危害。根据Hackman和Oldham（1976）的研究，自主性是五个核心工作特征/维度之一，它促使产生诸如体会到工作的意义、对工作结果承担责任以及了解工作活动结果的心理状态。自主性被定义为“工作为个人在安排工作和确定执行工作所使用的程序方面提供实质性自由、独立和自主的程度”。⁵³

⁵¹ 澳大利亚Comcare：侵扰性监控。

⁵² 澳大利亚 工作健康与安全（工作中心心理社会危害管理）操作规范，2024，第62页。

⁵³ J. R. Hackman & G. R. Oldham，《通过工作设计激发动力：理论检验》 组织行为与人类表现，第16卷（2）（1976年），第250-279页，第258页。

确实，人性化的工作条件的基本方面是“能够对自己工作不同方面施加影响的能力”。⁵⁴ 职业医学与心理学学科已显示出工作自主性（即在其三个维度上：行动范围、变动/创造力范围和决策自由度）更高与工作自主性发生频率更高的直接相关性。⁵⁵ 对工人健康产生积极影响。在一项关于心理社会工作条件的研究中，工作控制，即高决策自主性和任务多样性，也被确定为良好工作条件的一个决定因素。⁵⁶

根据世界卫生组织关于工作场所心理健康指南，“与工作控制相关（即在工作决策中拥有低权力）的因素与心理健康状况的症状相关；而更高的决策自由度对抑郁症状具有保护作用，更高的工作控制与减少情绪耗竭和倦怠相关——低工作控制与自杀风险增加以及与心理健康诊断相关的缺勤风险增加有关”。⁵⁷

那么，自动化如何影响人类的控制和自主性的问题随之而来。一项研究发现，在低至中等程度的自动化水平下，工人的表现最佳，而较高水平的自动化则与表现和工作量呈负相关。⁵⁸ 高水平自动化增加了心理负荷，对情境意识、控制感和任务多样性产生了不利影响。其他研究区分了执行任务的复杂程度，强调对执行过程的控制对于“需要高度专业技能的任务更为关键”，而简单且重复的任务则可以由完全自主的系统执行，而不会对工人的自主性产生负面影响。⁵⁹

最近，一项针对8000名员工的德国调查——“数字化与就业变化”（DiWaBe）发现，最低程度的自主性是“影响员工心理健康的一个重要因素，并代表着长期的心理疾病风险”。⁶⁰ 总体而言，这项研究的结果表明，当更多决策由技术做出时，工作条件恶化而不是改善。因此，研究指出，在人与机器的交互中，对技术的控制微乎其微甚至没有是职业自主性的威胁。因此，维护工人的自主权和工作控制权是这些研究的一个明确政策含义。

与过度监控的情况不同，关于自动化/基于AI的系统减少工作控制的心理社会风险的讨论非常少。诚然，工作控制作为因素可能涵盖了技术的影响，但职业心理学和医学领域并未包括与使用基于AI的系统相关的特定考虑，因此可能无法“充分解释与任务数字化相关的一些具体变化”。⁶¹ Since

⁵⁴ S. Niehaus, M. Hartwig, P. H. Rosen & S. Wischniewski, 职业安全与健康视角下的人类在控制和人工智能, 前人工智能 第5卷 (2022年)。

⁵⁵ Niehaus等人详细阐述了Eberhard Ulich关于工作自主性的理论框架，据Ulich所说，该框架包含三个组成部分：1) 行动范围，即“工作行动执行和时序组织中的自由度”；2) 变化范围/创造性的范围，描述为“工人设计其工作和程序的机会程度”；3) 决策自由度，理解为“员工决策权限和自主决定和界定工作任务的自由度”，在Niehaus等人中。 职业安全与健康视角下的人类在控制和人工智能, 第3页。

⁵⁶ T. Theorell, 《工作需求/控制/压力》。见：M.D. Gellman & J.R. Turner (编者), 行为医学百科全书 (纽约：斯普林格出版社, 2013年)。

⁵⁷ 世界卫生组织 工作场所心理健康指南, 2022。

⁵⁸ D. Kaber & M. Endsley, 自动化水平与自适应自动化对动态控制任务中人类表现、情境意识和工作负荷的影响 理论在人体工程学科学中的问题 e 5(2004), 113–153。

⁵⁹ Niehaus等人 职业安全与健康视角下的人类在控制和人工智能 第6页，还参考了L. Fréour、S. Pohl和A. Battistelli得出的类似结论：《数字技术如何改变工作特性：一项初步研究》 西班牙心理学杂志, 第24卷 (2021年)。

⁶⁰ Niehaus等人 职业安全与健康视角下的人类在控制和人工智能 第11页。

⁶¹ Niehaus等人 职业安全与健康视角下的人类在控制和人工智能 第11页。

看来人工智能系统的出现加剧了低工作控制的“风险”，需要进一步调查和政策分析。

过度收集数据和缺乏透明度

根据1997年国际劳工组织关于保护劳动者个人数据的实践守则规定，“雇主出于多种目的收集求职者和工人的个人数据：为了遵守法律；协助就业、培训和晋升的选择；确保个人安全、个人保障、质量控制、客户服务和财产保护”。此类信息作为业务运营的一部分是必需的，企业合法处理有关其员工敏感和个人的信息。

然而，随着高级数字技术的出现，如物联网设备（例如使用传感器捕捉信息的设备，如可穿戴设备）、机器学习应用和数据分析，以及通过对员工和工作流程的不断和全面监控而产生的大数据，被用于构建、测试和使用AI。正如一份报告所述：“数据可以用来‘创建一个反馈循环，其中AI系统的采用需要收集更多的员工数据’。”⁶² 这一大批数据的收集，其详尽程度和加工处理是前所未有的，远远超出了雇主应合法获取的信息范围。

此外，先进的数字技术在数据收集和处理方面缺乏透明度，放大了隐私风险，阻碍了数据主体对这些风险的认知。⁶³ 如所指出，“过度收集[数据]是对员工的一种过度曝光”。⁶⁴ 这可能导致人们感知到不公平和不信任。

⁶² 澳大利亚 [工作未来](#)，第45页。

⁶³ EUROFOUND [员工监控：监管的移动靶子](#)，2024。

⁶⁴ Rasmussen先生 [少即是多：过度收集移动设备数据的危险](#)，2024。

× 三项针对人工智能相关个人敏感数据的监管框架

全球工作领域的数字化转型跨越了各种国际劳工标准和不同层次的国家监管框架。以AI技术为基础的心理社会风险属于职业健康安全框架、规范数据保护和隐私的法律，以及工作关系框架的范畴。关键的国际劳工组织工具有1981年的职业安全与健康公约（第155号）、其伴随的建议No. 164、以及2019年暴力与骚扰公约（第190号）、其伴随的建议No. 206。

尽管这些工具可以为评估由基于人工智能技术带来的风险提供法律依据，但一个重要的问题是要问在职场部署基于人工智能的系统是否会动摇劳动法所基于的中立技术前提。⁶⁵ 任何尝试回答这个问题的努力都应该以这样一个前提开始：规范职场世界的法律框架是为了在传统的人与人之间的雇佣关系中创造工作条件。这一点得到了国际劳工组织宪章的证明，该宪章提倡并谈论“人道的劳动条件”，以及追求在“自由和尊严的条件”下的物质福利（即就业）。⁶⁶ ——这是国际劳工组织百年宣言在以人为未来的未来工作方法中反复重申的誓言。

对此，也有人可能认为职业安全与卫生立法构成了基于结果的法律框架，其中雇主的义务是在尽可能的范围内，提供一个安全的工作场所和不存在健康风险的工作环境。这些以结果为导向的职业安全卫生规则还伴随着基于流程的要求，如关于在职场引入和实施这些技术方面的咨询和信息义务。因此，职业安全卫生法规为解决由基于人工智能的系统带来的物理上和身心健康造成的损害提供了一个框架，前提是它们是基于国际劳工组织基本的职业安全卫生公约的，这些公约包括了与工作相关的身心健康各个方面。⁶⁷ 此外，职业安全卫生框架提供的预防性方法有助于更好地理解和应对与人工智能系统部署相关的风险，包括通过改进对这些技术创造或涉及的身心社会因素的映射和理解，并将它们整合到风险评估中。

这种现有职业健康安全（OSH）架构的明显适应性正受到越来越多的质疑，这促使人们呼吁，尤其是在欧洲背景下，采用“一种更新鲜、更广泛的职场健康安全定义，这种定义不能再与心理健康分开”。⁶⁸ 这次呼吁重新概念化的呼声进一步得到加强，因为各国“没有为职业安全与健康规定设定相同的具有法律约束力的共同标准和原则，这导致实际上的工人法律保护不平等”。⁶⁹ 这进一步说明职业安全与健康法规并没有始终如一地解决PSR问题，并且PSR仍然是全球范围内最不易被人理解的职场风险之一。

⁶⁵ 参见G. Le Mollé, [人工智能与人类尊严：当人类表现不佳成为法律要求](#)，2022。

⁶⁶ 国际劳工组织章程 序言和费城宣言，II。另请参阅 [国际劳工组织百年宣言：关于未来工作的未来](#) 其中序言提出了“人道的劳动条件”这一概念。

⁶⁷ 国际劳工组织百年宣言：关于未来工作的未来，P. II（函D）。

⁶⁸ 欧洲议会 《关于数字工作领域中精神健康的2022年7月5日决议》（2021/2098(INI)），第F段。

⁶⁹ 欧洲议会 《关于数字工作领域中精神健康的2022年7月5日决议》。

实际上，在许多司法管辖区，工作场所的健康与安全更多地被识别为物理安全而非健康。即使国家法律广泛地处理职业健康安全，涵盖心理健康和社会维度，但职业安全与健康报告（PSR）并不总是明确提及。关于工作心理健康决定因素的引用并未被具体化或整合到实施措施中。⁷⁰ 然而，即使在国家层面，心理社会危害被明确纳入立法并通过实际措施加以实施的情况下，针对人工智能系统对工作场所影响的研究一再强调，目前的劳动法规对“适应持续数字化转型”具有“技术中立性”，揭示了职业健康安全法规的空白和局限。⁷¹

人们对职业健康安全框架的充分性提出的新担忧，涉及人工智能技术所创建的“只在特定使用环境中逐渐显现的职业健康安全风险……[这些风险]动态涌现”。⁷² 这给现有的职业安全卫生（SbD，即设计安全）法规的“主流范式”带来了问题……因为它专注于设计阶段。⁷³ 在审视欧洲背景下的更多最近文件，例如职业安全与健康框架指令（FD），以及ISO 45-003（《心理社会风险管理工作指南》）时，一组研究人员指出，“所有这些工具都没有考虑到由持续部署人工智能而引发的之前未知职业安全与健康风险的动态出现：它们都遵循SbD传统，有效假定所有职业安全与健康风险均能在设计阶段得到解决和缓解”。⁷⁴ 不可预测因素似乎得到了人工智能系统开发者的证实。⁷⁵

学术辩论和监管举措开始探讨是否职业健康安全法律框架能够应对这些基本不同的情况，以及是否最终可能需要新的法规。⁷⁶ 例如，如前文所述，数字化工作场所中主要的风险之一是过度或侵入式的监控和监督。尽管这部分可能已被现有的心理社会风险管理框架所捕捉——通常在“组织文化”维度内——但它同时也突显了诸如隐私、数据保护等问题，这些问题在少数国家例外的情况下，并没有被一个全面的法律框架所综合解决。

在工作场所部署人工智能以及随之而来的过度数据收集是另一个令人担忧的领域。尽管数据保护在世界各地的环境中是一个熟悉的话题，但其监管充其量是局部的。在数字化工作世界中，“不仅工作（和工人）的管理与监督以新的数字方式出现，而且可能掌握在不同行为者手中，甚至可能是一个超出个人甚至雇主完全控制的自动化系统复杂体”。⁷⁷

因此，可以初步得出结论，目前，在没有全面监管框架的情况下，人工智能系统正在工作场所得到部署。防止在此背景下人工智能使用和操作带来的风险可能并不简单。如果有人提出指控，称

⁷⁰ 国际劳工组织 [通过职业安全与健康措施预防和工作中暴力与骚扰](#)，2024。

⁷¹ 澳大利亚 [工作未来](#)，2025。

⁷² A. Cefaliello et al. [确保算法管理对工人安全健康：新法律法规中应对心理社会风险](#)，第194页。

⁷³ Cefaliello等，第194页。

⁷⁴ Cefaliello等人，第201页。

⁷⁵ 例如，Dario Amodei，AI公司Anthropic的CEO和联合创始人承认：“……生成式AI系统是……” 长大 超过它们本身 建造 他们的内部机制是“涌现”的，而不是直接设计的。这有点像种植植物或细菌群落：我们设定了指导并塑造生长的高层次条件，但最终出现的具体结构是不可预测的，难以理解和解释……与生成式AI相关的许多风险和担忧最终都是这种不透明性的后果，如果模型是可解释的，这些问题将更容易解决。”——D. Amodei [紧急的可解释性](#)，2025。

⁷⁶ 澳大利亚 [工作未来](#)，第18页。

⁷⁷ F. Hendrickx [劳动者个人信息保护：基本准则](#)，国际劳工组织 WP 系列文献（2022），第10页。

歧视是基于人工智能操作，受到影响的个人是否有权上诉？受到影响的个人甚至是否知道人工智能正在被使用？如果他们知道了，责任在谁？是开发者还是部署者？那么雇主如何确定决策是由于算法偏见还是基于更大的数据集？

显然，鉴于数字技术及其不可预测性带来的根本不同情况，将OSH预防方法应用于基于AI的系统可能会具有挑战性，这可能会引发关于当前风险评估模型以及这些系统在就业环境中的治理、监督和管理的疑问。目前，讨论主要集中在基于AI的技术为改善工作场所安全和健康提供的机遇上，并且除少数例外，很少有人去描绘和理解这些技术的负面影响。这并不是说预防方法没有帮助，但这一事实表明，需要额外的法律阐述，以使规范性框架适应当代不断变化的工作世界。⁷⁸

在此提出，为了有效应对由人工智能引发的PSR问题，应通过包括劳动和就业法、平等与非歧视、职业健康与安全以及隐私和数据保护等几个领域的综合方式采取行动。国际劳工组织在最近的标准化活动中已经采取了这种监管方法。⁷⁹ 通过PSRs框架解决人工智能对心理健康的影响将是有效的，如果它能得到必要的监管解决方案的支持，以解决与工人数据保护、隐私相关的基本问题，并探讨“新”职业健康安全（OSH）权利的作用和地位，例如，有权对人工智能做出的决定提出上诉、有权断开连接等。为了为综合方法的构成打下基础，下文将对全球范围内试图解决工作场所人工智能影响的最新监管趋势进行概述。这些国家和地区的实践例子可以帮助我们确定人工智能监管的发展方向，以及职业健康安全问题得到解决的程度。

通用人工智能法规及其中的职业健康安全范围

尽管关于人工智能的指南、概念、伦理框架、政策和战略愿景众多，但尚未有太多具有约束力的法规。⁸⁰ 各国和地区机构开始设计和引入新的AI立法。总体而言，各国选择通过非约束性的概念和愿景来处理AI问题。这些内容将不在本次审查范围内。⁸¹ 为了本次分析的目的，将对现行法律和关于人工智能的草案法律进行审查，以分析职业安全健康（OSH）方面的特定内容。

只有少数国家制定了针对人工智能的具有法律约束力的文件（例如意大利、日本、韩国和越南）。在欧洲层面，欧盟已通过适用于所有欧盟成员国的AI法规。许多国家正在制定管理人工智能的法律（例如巴西、智利、哈萨克斯坦）。⁸² 法规方法对人工智能的适用以及其内容方面

⁷⁸ 更具体地说，“应对这些风险因素缓解策略包括确保数据隐私、让工人参与决策、增强数据解读中的人为责任，以及调整法律和政策框架。”在：EU-OSH，[数字技术在职场中的应用与心理社会风险](#)，2024。

⁷⁹ 例如，《2019年国际劳工组织暴力与骚扰公约》（第190号）倡导综合方法。同属此类，参见ILC，[平台经济中体面劳动委员会的结果：提交给大会通过的提案决议和结论](#)，ILC.113/记录号6A（2025）。

⁸⁰ 参见经合组织人工智能全球行动计划（OECD AI GPAI）。[政策观测站：国家人工智能政策和战略](#)。

⁸¹ 为了非洲联盟（AU），[大陆人工智能战略：利用人工智能促进非洲的发展和繁荣](#)，2024。

⁸² 样本国家基于以下两点：1）立法的可用性；2）可在公开来源中获取的立法草案的可用性。

全球职业安全与健康（OSH）差异很大。总体而言，尽管以人为中心的人工智能、以人为本的人工智能监管、道德和/或负责任的人工智能，以及权利、透明度、隐私和非歧视等概念在人工智能监管的背景下有所涉及，但在这些概念的内容以及具体到工作场所安全和健康方面，并没有明显的共同方法。⁸³

两点应予指出。首先，大多数关于人工智能的法规并未明确提及工人或雇主，而是确定了人工智能供应链中不同参与者的权利和义务。因此，在欧盟，参与者被称为部署者、提供者、授权代表、进口商、分销商、运营商、下游提供者和产品制造商。哈萨克斯坦的草案法律划定了所有者、持有者和用户的角色，越南的法律还额外将用户标明为执行方。韩国的立法也引入了“受影响者”的概念，指的是其生活、身体健康和基本权利因人工智能产品或服务而受到重大影响的人。⁸⁴ 这些术语与职场演员的世界并不完全相同，最终可能会对作为保护所有其他职场权利（包括职业安全与健康）基础权利的结社自由权造成挑战。

其次，现存的少数法规倾向于采用基于风险的方法来监管人工智能，全面的跨部门监管干预集中在以“产品安全为导向”的人工智能监管上。这种基于风险和安全的的人工智能监管方法并非基于雇主-雇员关系（即在职场环境中，而是将雇主（及其义务）作为代理人或部署者进行监管）。无论术语如何，人工智能法规确实适用于职场。在欧洲层面，欧盟人工智能法案中包含了对工作场所的引用，随同欧盟人工智能法案的文件也确认了人工智能法案适用于工作场所，声明如下：

□□ 人工智能系统在就业、工人管理和自雇准入中应用，特别是在人员招聘和选拔、影响工作关系条款的决策、工作关系的晋升和终止、基于个人行为、个人特质或特征分配任务以及监测或评估工作关系中的个人等方面，也应归类为高风险，因为这些系统可能对个人的未来职业前景、生计以及工人的权利产生显著影响。相关的工作关系合同，应有意义地涉及员工和通过平台提供服务的人员，如2021年委员会工作计划中所述。在整个招聘过程以及在评估、晋升或保留工作关系中的个人时，这些系统可能会延续历史上的歧视模式，例如针对女性、特定年龄群体、残疾人、具有特定种族或民族起源或性取向的人。用于监控这些人员表现和行为的人工智能系统也可能损害他们基本的数据保护和隐私权。⁸⁵

因此，人工智能系统根据它们对安全和健康的危险程度进行分类，或者在越南等地，根据诸如对人权或安全的影响程度等标准进行分类。⁸⁶

⁸³ 关于以人为中心的AI方法，请参阅意大利。法律第132号/2025年 2025年9月23日 并且越南，中华人民共和国人工智能法，第134号/2025/QH15，2025年12月10日。

⁸⁴ 韩国共和国，人工智能基本法 第20676号法律，2025年1月21日，第2条。

⁸⁵ 欧盟、人工智能法案 朗诵第57篇 关于第六条和附件三。

⁸⁶ 越南 中华人民共和国人工智能法，第134号/2025/QH15，2025年12月10日 第4条

某些工作场所中使用AI的情况在各个司法管辖区均有明确的规定，可能是高风险或不可接受的风险类别。在欧洲联盟，AI法案禁止使用AI系统在“工作场所和学校机构”领域推断自然人的情绪，但有例外“当使用AI系统的目的是为了医疗或安全原因将系统投入市场或市场时”。智利草案法律将AI系统的此类功能能力（即在工作场所推断自然人的情绪）定性为不可接受的AI风险。

潜意识操控系统，那些利用人们的脆弱性产生有害行为的系统；基于敏感个人数据的个体生物识别分类系统；通用的社会分类系统在各个法域的立法中被禁止，尽管具体措辞有所不同（欧盟AI法案、巴西和哈萨克斯坦的草案立法）。⁸⁷ 哈萨克斯坦的草案法律禁止完全自主的AI以及具有上述功能能力的AI。⁸⁸

同样，欧盟人工智能法案将人工智能在人的意识或目的之外部署的潜意识技术、故意操纵或欺骗性技术视为禁止行为。除此之外，还包括利用自然人或特定群体因年龄、残疾或特定社会或经济状况而具有的任何脆弱性的行为。这种担忧在许多法规中都很常见，因为其直接影响就业和更广泛的工作环境。情感识别技术被用于通过生物特征分析（例如面部微表情、语音模式与音调、头部和身体姿势、步态和脑部活动）来识别和解释员工的身心状态。⁸⁹ 虽然招聘是情绪识别技术应用的常见例子，但这些技术可以应用于整个雇佣关系的周期。因此，限制或禁止人工智能能力，如推断情绪、潜意识操控系统、基于敏感个人数据的个人生物识别分类系统、通用社会分类系统等，旨在保护工作环境中个人的隐私和自主权，并在一定程度上防止歧视和不公平行为。⁹⁰

然而，这些特定的法律限制在多大程度上涵盖了上述更广泛的身心社会因素，例如工作自主性、不受侵扰的监控或过度数据收集的工作环境，以及这些因素被涵盖的程度，提供的信息很少。例如，根据某些禁止的特征（包括种族、政治观点、工会会员资格、宗教和性取向）对个人进行生物识别分类的系统，在欧盟的AI法案中被禁止。⁹¹ 然而，生物识别技术可以在许多工作场景中使用——不仅用于招聘，例如通过“基于数据的候选人评估，例如自动化面试或心理测量评估”，还包括监控，如通过“生产力监控，例如通过键盘记录或运动传感器，或使用情感计算测量表现，集中度跟踪”以及解雇。⁹²

巴西的草案法案明确指出，在招聘、筛选、过滤候选人、关于晋升的决策等方面使用人工智能系统具有高风险。

⁸⁷ 参见例如越南的《中华人民共和国人工智能法》，第134号/2025/QH15，2025年12月10日 第7条。

⁸⁸ 哈萨克斯坦，《人工智能》草案法规，2025年1月。第18条。

⁸⁹ 参见欧盟。《人工智能法案》第18条

⁹⁰ TechDispatch 面部情绪识别 2021年（1），第2页。

⁹¹ 欧洲委员会 《关于禁止的人工智能实践委员会指南》（由法规（欧盟）2024/1689（人工智能法案）设立），2025。

⁹² 霍夫曼和马里内洛，《工作场所的生物识别技术：一种基于使用的分类法》（布鲁盖尔政策贡献2021）研究报告23/2021。

终止雇佣合同，任务分配及对受此类人工智能应用影响的就业、工人管理和自营职业领域的个人绩效和行为进行控制和评估。⁹³

欧盟《人工智能法案》的附件III明确列出高风险系统，并特别提及对工作场所的影响：

□□ (a) 旨在用于自然人的招聘或选拔的人工智能系统，特别是用于发布针对性工作广告、分析和筛选工作申请以及评估候选人。

□□ (b) 旨在用于决定影响工作关系条款、工作相关合同关系的晋升或终止、根据个人行为或个人特质或特征分配任务，以及监控和评估此类关系中的个人表现和行为的AI系统。⁹⁴

欧盟人工智能法案也包含了对人工智能高风险分类的豁免，因此该法案第6条第3款规定了从高风险分类规则中豁免的规定。根据附件三的定义，如果一个人工智能系统（如附件三所述）不构成对自然人的健康、安全或基本权利的重大危害风险，包括不实质性影响决策结果，则不应被视为高风险。此外，值得注意的是，作为其数字综合方案的一部分，欧盟委员会宣布推迟其实施人工智能法案的时间，将高风险人工智能系统的义务推迟至2027年12月，而不是原计划的2026年8月，以制定“创新友好的人工智能规则”。⁹⁵

意大利人工智能法案，作为欧盟人工智能法规的补充，也明确规定了职场中使用人工智能的问题，并将相关条款纳入标准劳动法概念之中。其第11条规定，职场中人工智能的使用必须安全、可靠和透明，不得侵犯人的尊严或个人数据的机密性。雇主必须通知雇员关于人工智能的使用。此外，该条款在劳动关系组织与管理中提出，人工智能确保了工人不可侵犯的权利，不受歧视。意大利立法还包含了一系列一般原则，包括人工智能系统必须遵守人类自主性和决策权、防止损害、可知性、透明度、可解释性等原则，以确保人类监督和干预。⁹⁶ 进一步研究的领域之一将是关于人类自主权和决策权在实际应用中的概念，以及这些概念与良好工作环境心理社会决定因素（如工作控制）之间的关系。

按照基于风险的方法，在符合一系列强制性规则的前提下，允许使用高风险人工智能系统，包括在高风险人工智能系统的整个生命周期内进行的“持续”和“迭代”风险评估。这需要部署者进行定期审查，以确保其有效性并尽量减少失败或故障，具体取决于人工智能的预期目的。事实上，智利草案法律与欧盟人工智能法案第9条相呼应，该法案要求建立风险管理机制，这被理解为持续

⁹³ 巴西第2338号法案草案，2023 第十七条

⁹⁴ 欧盟人工智能法案，2024年6月13日，第六条及附件三。

⁹⁵ 欧洲委员会 更简单的欧盟数字规则和新数字钱包将为企业节省数十亿资金并推动创新，新闻稿 布鲁塞尔，2025年11月19日。

⁹⁶ 意大利 法律第132号/2025年 2025年9月23日 第3条

迭代过程被计划并在高风险系统的整个生命周期中执行，需要定期进行系统性的审查和更新。⁹⁷

此过程包括以下步骤：(a) 识别和分析高风险AI系统在使用其预期目的时可能对健康、安全或基本权利构成的风险，以及合理可预见的风险；(b) 评估高风险AI系统在使用其预期目的和合理可预见滥用条件下的风险；(c) 根据《法案》第72条所指的市场后监控系统收集的数据分析，评估可能出现的其他风险；(d) 采取适当的、有针对性的风险管理措施，以应对根据(a)点确定的风险。⁹⁸

巴西的草案立法采取了一种略有不同的监管方法。它并不侧重于产品安全，尽管与欧盟类似采用基于风险的监管方法，但在强调尊重权利方面，它与其他所有监管倡议都截然不同。巴西的草案法律强调，在巴西使用AI系统必须基于包括社会权利在内，特别是“对人类劳动的尊重”。⁹⁹ 人工智能系统的开发应遵循福祉原则，包括保护劳工和工人。¹⁰⁰

巴西的草案立法还包括了与人工智能使用及其在工作场所影响相关的多项权利。因此，在草案法第5条中，受人工智能系统影响的人或团体（无论风险水平）享有以下权利：1）有权以可获取、免费且易于理解的方式了解其与人工智能系统的互动情况，包括关于交互自动化的性质（在极其有限的情况下除外）；2）有权保护隐私和个人数据；3）有权不受人工智能部署造成的非法或滥用歧视的影响，并有权纠正直接的或间接的非法或滥用歧视偏见。¹⁰¹ 4）解释权 5）对决策进行人工审核的权利¹⁰²

后者意味着一个受到AI系统影响的个人或群体有权了解系统所做的决定、推荐或预测；有权挑战和请求审查AI系统的决定、推荐或预测；以及在考虑到上下文、风险和技术发展状态的前提下，有权要求对决定进行人工审核。¹⁰² 这项解释权应通过简单、易于理解且适当的语言提供，以便在合理的时间内使人员理解相关决策或预测的结果，具体时间取决于人工智能系统的复杂性和涉及代理的数量。¹⁰³

尽管中国的监管框架¹⁰⁴ 被归类为（与贯穿性的主题即本简要审查的议题相对）领域性的，以下两条法规在工作权利方面值得提及。第一是中国《互联网信息服务算法推荐管理暂行规定》（2022年），其中第20条规定：“当算法推荐服务提供者对劳动者提供排班服务时，他们应当...”

⁹⁷ 智利，关于人工智能监管的草案立法，2024年5月。欧盟，[人工智能法案](#) 第9条

⁹⁸ 欧盟 [人工智能法案](#) 第9条

⁹⁹ 巴西 《2023年第2338号法案》 第2条

¹⁰⁰ 巴西 《2023年第2338号法案》 第3条

¹⁰¹ 巴西 法案第2338号，2023年 第五条

¹⁰² 布莱伊尔 法案第2338号，2023年 第六条

¹⁰³ 巴西 法案第2338号，2023年 第7条。

¹⁰⁴ 中国 《新一代人工智能发展规划》（2017），中国，[推荐算法管理规定](#)（2022年），中国，[临时管理生成式人工智能服务的措施](#)（2023）。

保护工人合法权益，如获取劳动报酬、休息和休假，建立和完善平台订单分配、薪酬结构及支付、工作时间、奖惩等相关的算法。¹⁰⁵ 在另一项名为《生成式人工智能服务管理暂行措施（2023）》的规定中，“当用户发现生成式人工智能服务不符合法律法规或本措施时，有权向相关部门投诉或举报”。¹⁰⁶

智利和哈萨克斯坦在一定程度上效仿巴西的做法，在各自的AI草案立法中。两国均纳入了规范AI系统的原则。这些包括隐私和数据治理、透明度和可解释性、多样性和非歧视、社会和环境福祉。¹⁰⁷ 在风险评估方面，智利草案法律将高风险人工智能系统定义为那些可能对健康、安全、基本权利或环境造成重大危害，以及损害消费者权利的系统。在欧洲背景下，劳动权利维度体现在欧盟人工智能法案第26条的第7款，主要关注工作场所使用人工智能的透明度。

在使用高危险等级人工智能系统前，雇主部署方应当告知工人代表及受影响的工人他们将接受高危险等级人工智能系统的使用。适用的情况下，此信息应按照《欧盟》及国内法律法规和关于通知工人及其代表的操作规范提供。

总的来说，基于这次粗略的比较审查，可以得出结论：目前监管人工智能的框架，除了一些例外，比如巴西的草案立法，并不是基于传统的劳动关系设计的。这些人工智能法规没有针对雇主、雇员直接提出，也没有明确针对工作场所中人工智能系统的应用进行调整。相反，如上所述，雇主被认定为供应链中的部署者和提供者，而在这一过程中，雇员几乎没有任何代理权或参与权。当人工智能法规涉及工人的健康和​​安全，并包括风险评估时，它们并没有充分解决与工作场所中人工智能系统使用相关的各种问题，如过度监控、工作自主权丧失、持续和过度收集数据等。这一差距可能部分通过劳动法在未来章节中将看到的方式得到填补。

国家针对基于人工智能的风险的职业安全健康相关法规应对措施

除了人工智能立法外，国家立法也针对人工智能对劳动和职业健康安全（OSH）法规中健康和​​安全的影响和风险进行了处理。这些监管发展特别关注工作场所监控的监管，特别是那些由技术支持的技术——这是一个以前未受国际标准监管的领域。与一般的OSH框架相比，以下监管示例并非技术中立，实际上似乎是为了应对雇佣关系的科技化，并试图在工作场所提供对身心健康的保障。

¹⁰⁵ 中国 《互联网信息服务算法推荐管理规定》 第二十条，非官方译本。

¹⁰⁶ 中国 临时管理生成式人工智能服务的措施 第十八条

¹⁰⁷ 智利，草案 人工智能规范立法 第4条至第11条 哈萨克斯坦，《人工智能》草案法规，2025年1月。

例如：瑞士 实施的法律——《就业法》（EmpO 3）第3号条例第26条¹⁰⁸ 那禁止使用监测或控制系统进行监控的 行为

（行为 在法语中）对工作场所员工进行监控。该法规还规定，在因其他原因需要监控或控制系统时，必须以不影响员工健康和正常活动能力、不使其处于持续监控之下为原则进行设计和安装。

该法规中的关键字是 行为 ，那么“这种行为”究竟指的是什么呢？因为，正如瑞士联邦数据保护与信息专员所明确指出，区分行为与表现往往是困难的——这两个概念常常是相互关联的。¹⁰⁹ 普遍认为后者，即绩效，需要评估和监控。专员指出，“在安全、绩效或效率的原因下，往往难以区分授权的监控与未经授权的一般行为监控”。因此，虽然可以允许用于验证绩效或输出的监控系统，但“通过持续、定期或抽样分析员工活动以监控其行为是被禁止的”。某些系统“根据其使用情况，可能违反监控员工行为的禁令”，可能包括“使用人工智能工具对基于员工的自动评估数据（视觉、运动、言语或沟通模式、心理结果）或监控系统员工的计算机或手机活动（间谍软件、活动追踪器、应用程序和网站日志、电子邮件内容扫描器、鼠标和键盘日志）”。¹¹⁰ 实际上，如果一个监控系统符合法律规定，它将由法院根据个案进行审查。

在澳大利亚（新南威尔士州），《2005年职场监控法》规范员工在工作场所的监控。此处区分了秘密监控和非秘密监控，除非得到秘密监控机构的授权，否则禁止秘密监控。监管框架随后详细说明了监控的条件。例如，规定在未提前（至少14天）以书面形式通知员工的情况下，不得开始监控员工。¹¹¹ 此类通知必须说明将要执行何种监控（摄像头、计算机或跟踪）、监控将如何进行、监控何时开始、监控是连续的还是间歇性的，以及监控是特定有限期限的还是持续的。¹¹² 该立法还规定了监控记录的使用和披露限制。¹¹³

法国是另一个实行广泛和侵入性监视限制的国家的例子。在法国，自2008年后，未经事先通知员工，不得引入监控（数据收集工具）。¹¹⁴ 因此，必须让员工们知晓监控系统的存在。本条规定应与法国劳动法典第L1121-1条款一起阅读，该条款规定，监督必须根据任务的性质予以正当理由，并与其目标相成比例。

在德国，2021年，《德国工作宪法法案》（第80条第3款、第90条第1款第3项和第95条第2a款）中纳入了与人工智能相关的修正案。第80条第3款

¹⁰⁸ 瑞士，《关于劳动法的第3号法令》（OLT 3）第5节 监工监督，第26条。

¹⁰⁹ 瑞士联邦数据保护和信息专员 工作场所监控系统说明。

¹¹⁰ 瑞士联邦数据保护和信息专员 工作场所监控系统说明。

¹¹¹ 澳大利亚（新南威尔士州），工作场所监控法2005年第47号，第10条。

¹¹² 澳大利亚（新南威尔士州），工作场所监控法2005年第47号，第10条。

¹¹³ 澳大利亚（新南威尔士州），工作场所监控法2005年第47号，第18条。

¹¹⁴ 法国，《劳动法典》，第L1222-4条。

第90条第1款规定了向工作委员会通报“包括人工智能使用在内的工作程序和操作”的义务，并为工作委员会就人工智能的引入或使用咨询专家提供了法律依据。¹¹⁵

工作混合形式的兴起（即远程办公、智能办公）¹¹⁶ 等等）已导致多个国家的立法调整，引发了对监控和监管的监管干预。在希腊，2021年通过了第4808/2021号法律，该法律在劳动法领域引入了重大改革，包括确保遵守2019年国际劳工组织《暴力与骚扰公约》（第190号）和远程工作规则，其中对员工个人数据等的重要规定有明确的规定。根据该法规，雇主应尊重员工隐私并以符合保护个人数据的方式进行绩效监控，禁止使用摄像头监控员工的绩效。¹¹⁷

在葡萄牙，2021年12月通过了第83/2021号法律，修改了远程工作制度。在第169条关于工作组织、管理和控制的内容中，规定远程工作中提供工作的指导和控制权力应依据其事先所知的程序行使……且应尊重其隐私和“雇主对提供工作的控制必须遵守比例和透明的原则，禁止在工作日通过图像或声音强制保持永久性连接”。¹¹⁸ 此外，第170条规定，远程工作制度意味着，对于雇主而言，包括以下特殊职责：在必要时，向雇员告知所有用于远程监控其活动的设备、程序和系统的特性和使用方法；在更改用于提供工作的设备和系统、分配的功能或合同活动的任何特性之前，以书面形式咨询工人。

在西班牙，于2021年通过了《工人宪章》统一文本的修正案，也称为…… *骑士法* 对雇主施加算法透明度的义务。¹¹⁹ 因此，根据修改后的文本，工会将有权从雇主/企业处获得有关“参数、规则和指示”（即“关于算法的逻辑和运行特性及其后果”）的信息，这些信息是算法或AI系统所依据的，并可能影响工作条件、就业机会的获取和维护，包括个人档案。¹²⁰

塞浦路斯关于远程工作组织框架的监管条例，于2023年通过，其第8(2)款规定，在实施用于监督、时间控制或评估员工绩效、效率和有效性的系统、技术、应用、措施或工具之前，雇主应进行数据保护影响评估，并与数据保护专员进行事先协商程序。

¹¹⁵ 德国 《企业宪法》（Betriebsverfassungsgesetz – BetrVG）。

¹¹⁶ 智能工作与远程办公相关，意味着利用信息和通信技术，使员工能在远离雇主场所的地方完成任务。

¹¹⁷ 希腊 法律第4808号 2021年6月19日，第67条（8）。

¹¹⁸ 葡萄牙 法律第83号/2021年 2021年12月6日，第169条，第4、5段。

¹¹⁹ 西班牙，《皇家法令第12/2021号》，2021年9月28日，修改《劳工法规总纲》，由《10月23日第2/2015号皇家立法条例》批准，以确保投身于数字平台交付领域人员的劳动权益。劳工和社会经济部指南明确，关于算法透明度义务的规定适用于使用无人工干预的自动决策影响工人的所有雇主。见MoL SE（西班牙）。

职场算法信息：企业使用职场算法信息义务指南与实际应用工具，2022年5月。

¹²⁰ 西班牙，2015年10月23日皇家立法法令2/2015，第64条。

规定还禁止使用摄像头或其他类似侵入性应用来监控员工的表现。¹²¹

2024年,《保加利亚劳动法》的修订包括了远程工作组织的相关规定。以下几方面值得提及。首先,该法规定,雇主必须提供关于法律要求和规则的信息,包括适用于远程工作的企业数据保护责任。¹²² 其次,如果需要在工作场所安装监控系统,必须事先获得工人或员工的书面同意。在这些情况下,应尊重他们的个人空间权利。¹²³ 第三,当远程工作的任务分配和报告通过信息系统进行时,雇主应向员工提供书面信息,说明系统中收集、处理和存储的工作相关数据的类型和数量。¹²⁴

第四,保加利亚劳动法典的新修订案规定了在远程工作组织中应用人工智能的适用。具体如下:

当使用针对远程工作算法管理的信息系统时,雇主应向工人或雇员提供书面信息关于男人决策网络的神经元

根据工人或雇员的书面要求,雇主或由其指定的官方有义务核查算法管理系统所做的决定,并通知工人或雇员最终决定。¹²⁵

“算法化管理信息系统”这一概念在保加利亚法律中被理解为“一种在分配、报告和控制员工工作时进行自动化决策的系统”。¹²⁶

这些由远程工作引发的监管规定,限制了侵入性的监控形式,涵盖了基于AI的系统,并且在职场数字化监管的前沿,与法国、瑞士和澳大利亚的例子在本质上有不同,因为它们仅限于远程工作。前者的应用范围更广,尽管有限,但也为无论工作如何组织的工作场所提供了保护。尽管如此,从对这些立法条文的审查中,我们可以得出几个原则:

- 当部署监测或控制系统时,必须根据任务性质证明其合理性,并且与预期目标成正比;
- 监测或控制系统的设计与安装不能对员工的健康产生不利影响;
- 在部署监测或控制系统时,应遵守一些条件,如事先通知和同意;
- 必须将监察的范围(例如:时长、程度等)告知员工。

¹²¹ 塞浦路斯 《2023年远程工作组织框架法规》(法律120(I)/2023) 第八条(第二款),非官方译本。

¹²² 保加利亚 劳动法 第107i条第(3)款和第(7)款,非官方翻译,以及修订后的英文版本。 编码 .

¹²³ 保加利亚 劳动法 第107i条(3)(8)。

¹²⁴ 保加利亚 劳动法 第107条h(10)。

¹²⁵ 保加利亚 劳动法 第107条h款(11和12)。

¹²⁶ 保加利亚 劳动法 第107条h款(11和12)。

监控/监视工具必须尊重隐私，并按照保护个人数据的原则使用。

系统内收集、处理和存储的工作相关数据的类型和数量等信息应以书面形式通知员工。

此类原则和措施将确保在职场监控背景下部署的基于AI的系统，受到旨在限制与侵入性监控和监督相关的PSR的程序性要求。

“新”职业健康安全权利的出现

工作场所中基于人工智能的系统监管引发了关于“权利”的讨论，例如了解人工智能系统信息的权利，以易于理解、免费且易于获取的方式，包括关于交互自动化的性质；隐私权利和个人数据保护；不受自动化决策歧视的权利，以及纠正直接或间接的不法或滥用歧视偏见的权利。¹²⁷ 解释权（当由基于AI的系统做出决策时）；最后，有权对决策进行人工审查。

这些权利类似于职业健康安全管理中固有的权利，即了解健康和安全问题的权利（包括雇主告知危害和风险、提供信息、指导、培训和教育的权利），参与可能影响健康和安全的决策的权利，以及拒绝可能影响工人及其他人健康和安全的作业的权利。

这些权利（获得充足的信息、指导与培训，以及对生命或健康构成紧迫和严重危险的作业环境中撤离）反映在国际劳工标准中，例如1981年《国际劳工组织关于职业安全与卫生公约》（第155号）。1988年《国际劳工组织关于建筑安全与卫生公约》（第167号）的第10条规定了关于参与权利的详细标准。它规定，国家法律或法规“应规定工人有权并有责任在任何工作场所就其对自己控制和影响的工作设备和方法的程度范围内，参与确保安全工作条件，并对可能影响安全和健康的采用的工作程序表达意见”。平台经济中关于体面工作的标准草案也包括了多项“算法透明度”规定，以及确保数字劳动平台健康与安全的标准。¹²⁸

这些标准可以作为制定与PSR相关安全和健康标准的基础。尽管这些对传统职业安全健康权利的新解释已适应工作场所算法现实的特定性，但它们尚未在国家级立法中取得显著进展。

在人工智能的约束性法规中，欧盟人工智能法案迄今为止在其第86条中纳入了个别决策解释的权利。此外，在欧洲背景下，通用数据保护条例相关并规定在第22条，数据主体有权“不受仅基于自动处理（包括档案整理）做出影响其或与其有类似重大影响的决策”。

¹²⁷ 巴西 法案第2338号，2023年 第五条

¹²⁸ 实现平台经济中的体面工作 草案公约和建议，ILC.114/报告V(4)。

他或她。¹²⁹ 然而，对于解释权的边界以及如何适用，尚未达成共识。

与此平行，还存在另一项职业健康安全权利，被称为“断开连接权”或R2D。¹³⁰ 近年来，将这项权利纳入劳动法规中引起了明显的兴趣。重要的是，某些政策文件将断开连接的权利指称为规范人工智能系统对健康和安全的法律工具箱中的方法之一。¹³¹ 不同国家在整合和实施法律策略以及监督法律执行方面采取了不同的法律策略。一些司法管辖区对其现有的工作时间和强制休息期的法律框架进行了修订，而另一些则在其软法工具中发展了这项权利，以工作场所安全和健康的一般权利为基础。¹³² 断开连接的权利是在立法文件中设计的，旨在部分保护工人的私人生活不受侵扰。它试图在私人生活和工作中建立平衡。问题仍然存在，即“断开连接的权利”，允许工人在工作之外的时间选择不参与工作通讯和监控，本身是否会成为直接的法律工具来保护或减轻有害人工智能系统的危害。¹³³ 一种方法将其视为允许与基于AI的持续工作场所监控断开连接，这是从工作场所心理社会健康角度关注的问题；这种做法将有利于更好地保护员工的健康和他们的隐私。

尽管这项权利确实有助于缓解对隐私的侵犯、限制侵入性监控或过度收集数据，但其在工作场所对健康和安全的影响，鉴于更多研究和证据尚待出现，似乎只是其他法律策略的补充，例如关于AI系统部署的信息知情权；参与和咨询AI部署的权利；隐私和个人数据保护的权利；基于自动化决策不受歧视的权利；解释的权利以及最后决策的人类审查权。未来工作环境中AI监管如何整合这些参数，以及更关键的是，如何应对机器与人类之间新的劳动分工，还有待观察。

¹²⁹ 根据欧盟通用数据保护条例第4条的规定，“‘个人数据’是指与已识别或可识别的自然人（‘数据主体’）相关的任何信息”。

¹³⁰ 参见ELI。《实施员工断开连接权利的指导原则》，2023。

¹³¹ 澳大利亚《工作未来》，2025。

¹³² 例如：澳大利亚，《公平工作法》2006年，第六部分；保加利亚，《劳动法典》，第154b条；比利时，《各种劳动规定法》，第16条；法国，《劳动法典》，第L2242-17条；希腊，《法律第4808号》，第67(10)条。

¹³³ 欧洲议会《应对职场中的人工智能风险：员工与算法》，2024。

✕ 结论

了解数字技术对工作场所健康和安全的影響，对制定工作场所人工智能安全使用政策至关重要。这份工作论文考察了现有（以及新兴的）关于技术变革对工作场所心理社会因素/景观影响的证据。研究发现，与工作场所人工智能使用相关的主要技术变革（例如，高级机器人、智能数字系统和基于AI的AM）与健康状况风险增加相关。更具体地说，论文列举了与工作场所人工智能系统和部署相关的经常引用的心理社会因素，并确定了一系列新兴的心理社会因素。这些包括侵入性监控、工作自主权和职场尊严的丧失、过度收集数据和职场透明度不足。

转向法律保护，该论文提出了对未来政策发展的建议，包括新的职业安全与健康权利。论文认为，需要一种综合方法来有效解决由人工智能引发的PSR。换句话说，应该在多个领域采取法律和政策行动，包括劳动和就业法、平等和非歧视法、职业安全与健康法以及隐私和数据保护法。正如研究中所强调的，在这方面已经取得了一些进展。例如，一些超国家和国家的劳动法规对工作组织的监管形式施加了限制，而其他司法管辖区已经开始具体设计关于工作场所监管的政策，无论工作是如何组织的（无论是远程还是在工作场所）。这些进展表明，涉及人工智能的侵入性监督和数据收集系统必须以不影响员工健康的方式设计和安装。当部署监控系统时，它们必须满足一系列条件，例如提前通知和工人的同意，以及必须让员工知道监督的范围（持续时间、范围等），并施加限制以尊重隐私并按照保护个人数据的方式使用。

尽管由过度收集数据产生的过度风险可能需要暂时对工人的隐私进行监管改进，但缺乏的是对更复杂的心理社会因素的监管，即工作自主权的丧失、工作尊严和职场透明度。这些问题引发了关于未来以人为中心的工作世界应该是什么样子的根本性问题。

参考资料

Aloisi, Antonio , Valerio De Stefano. 2022. 《你的老板是算法：人工智能、平台工作和劳动》。Bloomsbury出版社。

奇里科，弗朗西斯科。2017。“被遗忘的新兴心理社会风险领域。”职业健康杂志 59 (5) : 433-435。

Cox, Tom, 和 G. R. Oosterwijk. 2024. 工作场所的算法管理：北欧七个行业的算法技术影响案例研究。政策研究，十月。

De Tecco, C., B. Persechino, 和 S. Iavicoli. 2023. “职场变迁中的心理社会风险：从风险评估文化转向机会管理。” Med Lav 114 (2).

El-Helaly, M. 2024. “人工智能与职业健康安全：利与弊。”Med Lav 115 (2).

ELI. 2023. 执行劳动者断开连接权的指导原则。欧洲法学院。

欧盟-欧安安。2019。数字化与职业安全与健康。

2023-2025。健康职场运动：数字时代的安全与健康工作。

_____. 2024. 工作中的数字技术及心理社会风险：对职业安全与健康的证据与启示。

欧洲议会研究服务局 (EPRS) . 2022. 德斯蒂法诺，瓦莱里奥，和马蒂亚斯·沃特斯. 工作场所管理与评估中的AI和数字工具：欧盟法律框架评估。

欧洲议会. 2022.关于数字工作环境中心理健康问题的决议 (2021/2098(INI)) 。

Fréour, Léa, 和 Battistelli, Adalgisa. 2021. “数字技术如何改变工作特征：一项初步研究。” 西班牙心理学杂志 24。

Hackman, J. Richard , Oldham, Greg R. 1976. 《通过工作设计激发动力：理论测试》。组织行为与人类表现 16 (2) : 250-279。

亨德里克斯，F. 劳工个人数据保护：一般原则，国际劳工组织工作文件，2022年。

霍华德，J. 和舒尔特，P. 2024. “管理职场AI风险与工作未来。”美国工业医学杂志，1-14。

国际劳工组织 (ILO) . 1981. 职业安全与健康建议 (第164号) . 日内瓦：国际劳工组织

_____. 1984. 工作场所的心理社会因素：认识与控制。国际劳工组织/世界卫生组织职业健康联合委员会第九次会议报告，日内瓦，1984年9月18-24日。职业安全卫生系列第56号。

_____. 2025. 数字化带来的挑战与机遇。GB.350/HL/1(修订1)。

_____. 2025. 重新定义健康与安全：工作中人工智能与数字化的作用，第6页。

Jarota, Maciej. 2023. “工作流程中的人工智能：从职业健康与安全视角反思欧洲联盟人工智能法规的提案。”计算机法律与安全评论 49。

Jung, Kyungran, Yang Jae-Suk. 2025. 《缓解人机协作中的安全挑战：人类能力的作用》.科技预测与社会变革213期。

Le Moli, Ginevra. 2022年. “人工智能与人类尊严：当人类表现不佳成为法律要求。”8月。

Leka, Stavroula, Jain, Aditya, Iavicoli, Sergio, 和 Di Tecco, Cristina. 2015. “欧洲联盟工作场所心理社会风险和心理健康政策背景评估：成就、挑战与未来。”生物医学研究国际期刊。

李亚, M., 卡塔利娜-罗梅罗, C., 卡尔沃-博纳乔, E., 和贝纳维德斯, F. G. 2016. “工作中暴露于心理社会风险因素与职业伤害发生率：西班牙队列研究。”职业与环境医学杂志 58 (3) : 282-286。

麻省理工学院斯隆管理评论。2023. “算法管理：人工智能在劳动力管理中的作用。”四月。

Niehaus, S., Hartwig, M., Rosen, P. H., 和 Wischniewski, S. 2022. “商业中的AI。”人工智能前沿 5 (2022年7月) : 第3、6、7、11节。

经合组织 (OECD) , 2025 , 《人工智能论文：工作场所的算法管理——经合组织雇主调查的新证据》。二月。

澳大利亚议会，就业、教育和培训常设委员会。2025年。工作未来。1月。

Rani, Uma, Andrea Pesole, 以及 Iñigo Gonzalez Vazquez. 2024. 《常规工作场所的算法管理实践：物流和医疗保健案例研究》。2月19日。

Theorell, T. 2013. “工作需求/控制/压力。”收录于M.D. Gellman和J.R. Turner编辑的《行为医学百科全书》中。Springer, 纽约。

Varma, Arup, Vijay Pereira 和 Parth Patel. 2024. “人工智能与绩效管理。”组织动力学 53 (1) 。

Varma, V. Pereira & P. Patel, 人工智能与绩效管理，组织动力学，第53卷，第1期，2024年。

世界卫生组织. 2022. 工作场所心理健康指南

✗ 推进社会公正，促进体面劳动

国际劳工组织是联合国负责世界劳动领域的机构。我们汇聚政府、雇主和工人，旨在改善所有人的工作生活，通过创造就业、工作权利、社会保障和社会对话，推动以人为中心的工作未来。

联系信息

研究部（研究）

国际劳工组织
Route des Morillons 4
1211 日内瓦 22
瑞士
T +41 22 799 6530
research@ilo.org
www.ilo.org/research



ISBN 9789220434086

9HSTC

9 789220 434086