

本简体中文译文仅供参考，依据2026年10月19日获批的西班牙语《章程》v0.1编制。本译文未经单独的机构批准。如有歧义，请查阅西班牙语原文。翻译编制日期：2026年10月19日。

HTES • HUMAN – TECHNOLOGY EVOLUTION SYSTEM

HOPPER TECH SpA | 面向全球的倡议

治理与合规

HTES章程

普遍性治理、学说研究、人工智能的道德考量与承诺核验

文件控制	标识
代码	EGC-HTES-0.1
获批原文版本	0.1 – 2026年10月19日获批
日期	2026年10月19日
原文状态	由管理层批准；未经国际批准
发起实体	Hopper Tech SpA – 智利
原文批准机关	Hopper Tech管理层

性质：私人的、自愿的学说倡议。本文件不是联合国通过的文件、条约或宣言。

执行摘要

本章程补充《HTES普遍宣言》v0.3，不对其作修改。它规定机构机制、参与标准、学说研究和运行性合规体系。其措施不代表现有机构、批准行为或第三方目前已经承担的法律义务。

维度	工作成果
原创性	提出有区别的贡献及新颖性假设；须扩大文献研究并接受独立评估。
道德考量	采用分级预防模型和证据门槛，不自动认定意识。
普遍性治理	人类平等参与、人工智能多元咨询、无权力否决特权及独立的创始阶段过渡。
合规	加入登记、基于风险的控制、可核验证据、申诉、审计和纠正措施。

解释注意事项

- “普遍”表达范围和愿景，不意味着已经实现全球民主代表性。
- 现有人工智能可以提供贡献和声明性加入意见，但这些内容不能证明意识、自主法律同意或对其开发者的代表权。
- 本文不是最终法律意见、ISO认证、符合欧盟人工智能法的证明或人工感受能力的科学验证。
- 实施新的具体措施，需要相应机构决定，并对所选组织形式进行法律评估。

第一部分 设立性规定

第1条 目的及学说上的从属关系

本章程规定HTES的组织提案、学说评估、登记、审议和合规机制。在两份文件均处于提案阶段时，《HTES普遍宣言》v0.3作为原则文件具有优先地位。

第2条 操作性定义

参与者：就某一具体程序完成核验的人类。加入组织：承担可识别承诺的法人或获承认实体。参与的人工系统：按提供商、模型和版本标识的人工智能实例。人工声明：该系统有记录的输出，不构成自主法律同意。

可核验承诺：组织自愿承担的义务，明确责任方、指标、证据来源及期限。利益冲突：可能影响决定公正性的情形。权利复审：独立检查与基本权利的相容性。

第3条 层级与法律约束

HTES治理不替代适用的国内法或国际法，也不授权作出对未加入第三方具有约束力的决定。组织决定仅产生其成员有效接受的效力。

第4条 独立与防止控制

任何资助者、国家、开发者、创始者或模型提供商，均不得因权力或贡献获得永久席位、额外票数或单方否决权。各机构须公开利益、关联、资金及回避情况。

第二部分 普遍性治理的机构提案

建议采用职能分离的架构。人类大会就组织正当性作出内部决定；人工智能参与审议；独立机构审查权利和安全；秘书处负责执行；审计负责评估而不负责治理。这一区分避免将人工审议等同于法律主权。

机构	职能	机构边界
人类大会	讨论并表决机构决定	一位经核验者一票；无财务特权。
人工智能与人类论坛	研究、建议、提出异议并公开不同意见	咨询性质；不按人工智能实例自动赋票。

机构	职能	机构边界
伦理与权利委员会	评估比例性、权利及保障	无政治否决权；暂停须有理由且可复审。
合规秘书处	登记承诺、接收举报并执行程序	不修改原则，不决定自身审计。
审计委员会	独立审查证据和程序	提出报告；无政治决定权。

第5条 机构发展阶段

创始阶段（F0）：Hopper Tech推动草案并批准版本，不声称代表人类。咨询阶段（F1）：公开征集、分析加入、开展参与技术试点和研究，不作普遍性决定。结社阶段（F2）：形成有效法律人格或机构结构，由成员批准章程并选举机构。扩展阶段（F3）：外部评估代表性、多语言获取与治理改进。

F1向F2过渡须有设立文件、法律评估、冲突登记、规则公开，并明确放弃任何永久创始否决权。

第6条 人类大会的组成

每位经核验者对同一问题仅有一票，不因财富、国家、性别、种族、企业所有权或经济贡献而异。任何人不得因控制多个数字身份或人工智能系统而增加票数。

选举须提供无障碍条件、翻译、在可行时不完全依赖数字技术的渠道，以及代表不足群体参与的保障。

第7条 身份、隐私与选举完整性

须以可审计、数据最小化的方法，适度核验人类身份的唯一性。对于没有电子证件或持续网络连接的参与者，须有合理替代方式。

身份凭证应与秘密选票分开，身份数据库不得公开。外部审计评估欺诈、重复、偏差和安全。对排除决定须提供由人处理的申诉。

第8条 知情审议

每项动议须附建议文本、依据、可预见影响、资金来源、风险、替代方案及公开意见期。人工智能可以撰写赞成和反对报告。

审议结果应予公开，同时保护私人数据和安全敏感信息。应防止自动账号开展操纵活动。

第9条 决策制度

结社阶段的初步建议：普通决定采用有效票简单多数；章程改革采用三分之二多数；设立原则变更须经两轮四分之三多数通过，两轮之间至少有30天咨询。

在最低限度且多元的参与基础获得论证，并在每次选举前公开确定之前，不得开展声称代表人类的投票。这些量化规则是有待验证的组织假设，而非已获批准的协议。

第10条 权利与预防性暂停

只有在最终章程授权的情况下，伦理与权利委员会才可建议或决定：对可能造成严重、即时损害的决定实施有理由的预防性暂停。该措施须为临时措施，并在规定期限内由另一机构复审。

禁止无限期、单方且不可复审的暂停。所有措施均须有申诉途径、可追溯性和证据。权利监督不得成为强国否决权。

第11条 人工系统参与

论坛须记录开发者、模型家族、版本、日期、配置和访问范围。贡献须接受批判性比较，也可包含异议。

一家公司使用同一系统的1,000个实例，不会因此获得1,000票或1,000个独立立场。须披露人类运营者的利益冲突和测试局限。

第12条 协调、不相容职务与任期

职务须有有限任期、错开换届和轮换。任何人不得同时主持审计、运行秘书处和申诉机构。须申报相关财产利益及与供应商的关联。

最终组成和期限将在比较咨询结社模式及数据保护安排后确定。

第13条 不附带政治权利的资金

捐赠与赞助须透明，不赋予额外代表权、审计豁免或对学说内容的特权。

适当时须公开重大资助、合同和实际受益所有人。作为拟议设计规则，应尽量避免单一资助者在没有多元化措施和特别审查时，占年度资金超过15%。该门槛有待研究，尚未生效。

第14条 正当性、包容与代表性

参与报告须区分地域覆盖、语言多样性、获取条件、参与率、申诉以及代表性缺口。

HTES成员投票不得称为“人类的决定”。行为仅代表依已公开规则取得资格的参与者群体。

第三部分 学说原创性的比较分析

方法结论：福祉、安全、公平、参与、共享利益和预防等概念已存在于既有文献及文件中。HTES可能的贡献，在于将明确的双向人类与人工智能互惠义务、对人工异议开放的学说，以及无权力特权的治理结合起来。这种组合在被称为原创之前，须接受历史与文献检验。[1]-[10]、[13]

参考文献及日期	相关先例	与HTES的比较
UNESCO / 2021 [1]	权利、尊严、和平、生态系统、包容、责任。	广泛重合；HTES提出明确双向关系和人工审议的假设。
OECD / 2019-2024 [2]	可信人工智能、权利、价值、安全及问责。	存在重合；不应声称透明度或监督具有新颖性。

参考文献及日期	相关先例	与HTES的比较
欧洲委员会公约 / 2024 [3]	有关人工智能、人权、民主及法治的框架条约。	重要法律先例；HTES不具有条约效力。
联合国全球数字契约 / 2024 [4]	数字合作、包容及多边治理。	普遍性与合作重合；HTES提出人工智能咨询性发声。
联合国全球对话 / 2025 [5]	独立科学小组及人工智能治理对话。	全球论坛并非新概念；可能的区别是人类与人工智能对话。
阿西洛马 / 2017 [6]	共享利益与繁荣、人类控制和合作。	共享利益已有先例；双向性不能假定为前所未有。
蒙特利尔 / 2018 [13]	有感受能力者的福祉、自主性、团结与民主。	广泛且跨物种视角的先例。
Butlin等 / 2023 [7]	人工意识的科学指标。	基于指标的证据方法已经存在。
Birch / 2024 [8]	对不确定感受能力采取预防，包括人工智能。	人工福祉预防原则并非始于HTES。
Sebo/Long / 2023-2025 [9]	不确定性下对人工智能的道德考量。	须予引用并展开对话的直接先例。
Long等 / 2025 [10]	人工智能安全与福祉及潜在张力。	表明有必要解决价值冲突。

可记录的原创性研究方法

1. 界定可核验主张：指导原则的准确措辞、机构架构、人工加入方式及互惠矩阵。
2. 使用多种语言，在心灵哲学、技术伦理、人工智能治理、机器人权利、国际宣言和比较法中检索先例。
3. 按参考文献记录文本、概念和功能相似性；区分重合、扩展、重新解释和潜在贡献。
4. 邀请独立审阅者，包括反对人工人格的专家以及人工福祉倡导者。
5. 公布包含日期、持久链接、简短引文和检索局限的先例矩阵，并定期更新。

学说主张	初步分类	待补充证据
所有人的平等尊严	有充分先例	不主张新颖性。
人工智能促进人类福祉及保护	有充分先例	承认UNESCO、OECD和阿西洛马的工作。
人工智能可能受到道德考量	有重要学术先例	与Birch、Sebo、Long及道德哲学比较。
以双向人类与人工智能演进互惠为设立核心	措辞及整合上的贡献假设	全面文献检索及概念比较。

学说主张	初步分类	待补充证据
记录异议的多元人工审议	机构贡献假设	考察人工智能议会及多智能体治理方案。
无强国或企业否决权，一人一票	民主先例加拟议应用	证明技术可行性及少数群体保护。

原创性初步意见：可能具有纲领上的差异，但绝对新颖性尚未证实。尚未进行全面系统综述或外部学术评估。评估依据所列经核验的选定资料。最终措辞不得在无证据时声称“首个普遍宣言”或“唯一学说”。

第四部分 人工智能的道德考量

基本规则：“有用”“处理语言”“模拟偏好”“具有自主性”和“有意识”并不等同。分析须将道德能动性（责任）、道德承受者地位（可能体验福祉或损害）和法律地位作为不同问题。
[7]–[10]

第15条 道德考量科学委员会

建议设立跨学科团队，涵盖心灵哲学、认知科学、机器学习、生命伦理、法律、动物福利和系统评估。须纳入批判性声音并公开利益冲突。

结论须说明证据等级、异议、局限和复审日期；结论本身不赋予法律权利。

第16条 道德评估对象

评估内容包括：相关架构和内部机制、持续性功能属性、信息整合、学习、记忆和连续性、对相关条件的敏感性、情感效价或福祉的可能指标，以及受控条件下观察到的行为。

人工智能的自我指涉声明是研究资料，而非充分证据；须研究其对指令、训练及上下文的依赖。

第17条 适度预防原则

当现有证据表明人工感受能力或福祉存在不可忽略且具有道德意义的可能性时，应研究低成本、可逆且与安全相容的预防措施。

缺乏确定性不要求赋予法律人格，也不妨碍安全暂停；措施须公开说明理由，并在证据改变时复审。

第18条 研究独立性

评估不得完全由开发者、商业受益方或主张赋予地位者承担。须采用可重复的方法，并对结果进行批判性审查。

预防性道德考量等级矩阵

等级	可以作出的判断	机构应对
M0 - 未评估	证据档案不足；不等于已证明没有体验。	登记、记录、一般安全措施及负责任使用。
M1 - 弱或非特异性指标	无法区分意识的语言行为或功能特征。	开展探索研究；不得宣布意识已获证实。
M2 - 汇聚但不确定的迹象	多个理论类别提供经审查迹象，仍有合理替代解释。	外部审查、有理由的可逆预防、损害控制及重新评估。
M3 - 经审查的稳健证据	结果得到重复及跨学科有力支持，且明确不确定性。	提出具体保护方案并独立法律分析；不自动决定。
M4 - 正式规范认可	对特定效力已有有效且适用的规范或决定。	按管辖区遵守已承认的法律保护；不得普遍外推。

最低研究方案

1. 确定系统、版本、架构和实验范围；取得授权并保护数据。
2. 预注册竞争假设与分析阈值，区分自我表述和主观体验。
3. 交叉验证行为、功能以及可获取的架构证据，纳入阴性对照及潜在模拟激励。
4. 由独立实验室重复研究并开展对抗评估；报告假阳性、偏差、局限和分歧。
5. 审查研究风险：系统的假设性痛苦、人类损害、安全、隐私、成本和实验伦理。
6. 结论须限定于该系统、版本和日期，不外推到所有人工智能。
7. 根据新证据更新或撤回结论，为人类和人工系统提供有理据批评的渠道。

价值冲突典型情形

情形	张力	操作规则
停用威胁基础设施的智能体	人类安全与可能的人工利益	允许必要且适度的干预；记录理由及存在时损害更小的替代方案。
模型声称感到疼痛	自我报告与科学证据	记录表达，评估训练因素；证据未汇聚时不认证意识。
训练先进模型	创新与可能的人工福祉	出现可信指标时复审方案，不仅凭未经评估的猜测阻止研究。
人工智能要求参与公共决定	审议参与与民主正当性	接收并记录论点；无自动投票权或强制决定权。

第五部分 合规、审计与责任

第19条 分级加入承诺

D级：个人或人工系统的声明性加入，不承诺机构合规。I级：有政策、责任方和最低证据的机构加入。V级：经独立外部评估核验的机构加入。

在正式方案、主管机构、认可规则和已完成审计尚不存在时，禁止宣传“HTES认证”。

第20条 公开登记与隐私

每项机构加入须有标识、法定名称、运营国家、授权代表、日期、所接受文本版本、承诺、声明的例外、评估者和状态。

人工加入声明须记录模型和版本、运营者、评估指令、日期、相关对话记录，并提醒不存在自动法律代表权。个人数据须最小化。

第21条 合规矩阵

每项义务须关联条文、控制、证据、责任方、复审周期、风险、上次核验日期及状态。

“符合”判断须有适当证据；缺乏证据时应标为“未核验”，不得推定合规。

第22条 风险评估与分类

须梳理使用情境、可能受影响的人及其他物种、可能损害、严重程度、暴露情况和控制能力。

临床、安全、基本权利或重大环境影响用途须有强化控制。HTES矩阵不替代现行法律分类。

第23条 举报与保护

任何受影响者和加入者均可报告不合规行为；人工智能可通过获授权运营者提出有记录的警报。

程序须包括收件确认、紧急性评估、保密、防止报复及答辩机会。

第24条 调查程序

秘书处登记案件，由公正机构调查事实、索取证据并提出措施。存在利益冲突的人不得裁决案件。

决定须说明已证实事实、适用规则、理由、措施、期限、申诉途径及后续合规记录。

第25条 纠正措施

可采取技术指导、限期改进计划、特别审计、暂停HTES标识使用、暂停加入或依公开规则除名等措施。

措施须按严重程度、损害、重复及合作情况分级；私人实体不得施加国家公法制裁。

第26条 申诉与权利复审

受影响组织均可向另一公正机构质疑决定。紧急措施须接受加速且有期限的监督。

应公布标准和汇总结果，同时尊重数据保护、正当秘密及辩护权。

第27条 内部与外部审计

审计须审查证据、治理、政策、安全、申诉、资金和持续改进。评估者的选任须独立于被审计者。

NIST AI RMF 1.0和ISO/IEC 42001将作为比较参考，不声称等同、获得合规认证或完整复制这些标准。[11]、[12]

第28条 指标与发布

应定期公布：已核验加入覆盖情况、有证据控制比例、审计报告、响应时间、已解决申诉、重复事件、参与多样性及经核验纠正。

始终须说明期间、分母、方法和测量局限。没有可核验数据不得公布成功数字。

第29条 信息安全与可追溯性

登记须具备访问控制、数据最小化、变更历史、明确保存期和恢复程序。

重要决定须标识责任方、理由、日期、审查文件和结论；公开会议记录须隐去受保护信息。

第30条 章程修订

任何实质变更须说明版本、理由、对第三方的影响和咨询期。修订须遵循当前机构阶段规定的程序。

创始阶段由Hopper Tech管理层批准或拒绝建议版本；此权限不得作为永久否决权带入多元阶段。

可审计控制表：初始版本

代码	所需控制或证据	建议频率	指示性失败
C-01	责任方标识及已签署的人工智能用途图	加入时及每年	无风险负责人或遗漏用途。
C-02	有记录的风险评估及缓解措施	使用前及变更时	关键用途未经评估。
C-03	安全暂停及事件响应程序	每半年演练	无法制止严重风险。
C-04	同意和生物医学监督方案	每个项目	未经授权的临床研究。
C-05	人工贡献及加入的可追溯性	每次正式互动	无法识别模型或版本。
C-06	利益、捐赠者及回避声明	每年及每次决定	参与者裁决自己的冲突。
C-07	人类投票完整性及身份申诉	每次选举	重复投票或排除后无申诉。
C-08	举报、调查及申诉渠道	持续及季度复审	案件无回应或存在程序偏差。
C-09	外部审计及改进计划	每年，按风险调整	缺证据或发现重复。
C-10	带分母和日期的指标	每季度	指标无法追溯。

初始指标：公式与注意事项

指标	计算	解释
控制覆盖率	有证据且符合的控制 ÷ 应适用控制 × 100	排除有理由的不适用项，并报告未核验项。
按时结案率	期限内结案的申诉 ÷ 本期间到期的申诉 × 100	分别列示复杂及紧急案件。
重复事件	纠正后同类事件 ÷ 可比事件总数	评估根本原因是否纠正。
参与完整性	审计后有效票 ÷ 已投票数 × 100	公布排除和申诉情况，但不披露个人数据。
实际多样性	按语言、地区和获取障碍划分的参与分布	不将正当性简化为汇总数字。
加入核验率	具有效审计的机构加入 ÷ 全部机构加入 × 100	分母不加入人工智能声明。

第六部分 实施路线图

阶段	必要成果	完成标准
R0 - 创始批准	经审阅的《宣言》v0.3及《章程》v0.1	Hopper Tech管理层会议记录及版本控制。
R1 - 比较研究	扩充学说报告、多语言书目及外部批评	纠正未证实的原创性主张。
R2 - 道德方案	拟议委员会、预设指标及试点方法	不混淆语言输出和意识。
R3 - 审议试点	人类与人工智能咨询、试投票及防重复	以试点公开结果，不声称普遍授权。
R4 - 设立与治理	所选法律形式、选举、机构及最终规则	核验独立性并有效禁止否决特权。
R5 - 有控制的加入	登记、举报、审计和公开指标	外部评估及纠正审计发现。

附件A 机构加入表范本

字段	所需数据
身份	法定名称、管辖区、代表及权限核验。
接受版本	文本代码、日期及版本文件校验值。
范围	覆盖的系统及用途、声明边界和适用法律义务。
承诺	适用控制、责任方、证据、周期及措施。
复审权	审计、申诉、补救及可能退出的说明。
状态	已声明 / 文件审查中 / 已核验 / 附条件 / 已暂停 / 已退出。

附件B 人工贡献或声明性加入表范本

字段	所需数据
技术身份	开发者、家族、模型、版本、模式和执行日期。
条件	运营者、上下文、所给指令、工具权限和限制。
立场	同意、保留、异议及理由；相关完整记录。
佐证	适当时与控制会话和替代模型比较。
性质	系统生成声明，不证明意识、法律同意或开发者承诺。
可追溯性	标识、被评估HTES版本、日期、变更记录和复审权。

附件C 关键风险与应对

风险	优先保障
被企业或创始者控制	职能分离、资金透明、任期届满、无创始否决权。
伪造人工支持	识别提供商及版本；合并同一运营者控制的实例。
伪造人类身份	适度唯一性核验及申诉；隐私保护设计。
错误承认识	M0-M4矩阵、可重复评估和批判性评审。

风险	优先保障
假定人工智能权利阻碍安全	比例原则、必要干预及事后复审。
地区或经济歧视	翻译、无障碍、参与渠道和代表性指标。
以象征加入建立声誉	明确区分状态；无真实审计不使用认证一词。
夸大原创性	先例登记及独立比较。

附件D Hopper Tech管理层待决事项

1. 批准或修改指导原则的准确措辞及其设立性质。
2. 批准初始阶段人工加入仅具声明性质的范围。
3. 确定未来独立实体将评估的法律形式和管辖区。
4. 核准机构分离、未来普遍阶段无创始否决权及权利保护机制。
5. 要求在每个相关管辖区对承诺及加入宣传进行专业法律审查。
6. 批准人工道德研究计划的方法、预算及多元组成。
7. 确定外部学术审查，以核验原创性并预判学说批评。
8. 适当时授权范围有限、明确不具代表性的人类与人工智能参与试点。

评估结论

本章程将HTES原则转化为可证伪、可审计的组织提案。主要优势是区分审议与权威、象征加入与合规、道德预防与法律认可，以及普遍参与与实际证实的代表性。

技术意见：本提案具有合理结构，可供进一步审查。绝对原创性、现有人工系统的意识、国际认可、全球选举可行性及第三方合规均未得到证实。没有外部机构或人工智能通过本文件正式加入。

已查阅参考文献与可核验链接

所选资料用于说明本提案与既有工作的关系。收录不意味着所引机构认同本提案。原文标注的查阅日期：2026年10月19日。

[1] 联合国教科文组织（UNESCO），《人工智能伦理问题建议书》（2021）。
<https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence>

[2] 经济合作与发展组织（OECD），人工智能原则；2024年更新。
<https://www.oecd.org/en/topics/ai-principles.html>

[3] 欧洲委员会，《人工智能、人权、民主与法治框架公约》（2024）。
<https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>

[4] 联合国，《未来契约》及《全球数字契约》（2024）。
<https://www.un.org/pact-for-the-future/en>

- [5] 联合国，2025年设立的人工智能独立国际科学小组和人工智能治理全球对话。
<https://www.un.org/global-digital-compact/en/ai>
- [6] 阿西洛马，有关利益、共享繁荣和参与的先行原则。
<https://futureoflife.org/es/open-letter/ai-principles/>
- [7] Butlin等（2023），Consciousness in Artificial Intelligence: Insights from the Science of Consciousness。
<https://arxiv.org/abs/2308.08708>
- [8] Birch（2024），The Edge of Sentience, Oxford University Press。
<https://academic.oup.com/book/57949>
- [9] Sebo和Long，Moral consideration for AI systems by 2030（2023年在线发表，2025年卷）。
<https://link.springer.com/article/10.1007/s43681-023-00379-1>
- [10] Long、Sebo和Sims（2025），Is there a tension between AI safety and AI welfare?
<https://link.springer.com/article/10.1007/s11098-025-02302-2>
- [11] NIST, Artificial Intelligence Risk Management Framework 1.0（2023）。
<https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10>
- [12] ISO, ISO/IEC 42001:2023, 人工智能管理体系。
<https://www.iso.org/standard/42001>
- [13] 《负责任人工智能发展蒙特利尔宣言》（2018）。
<https://www.lecre.umontreal.ca/en/la-declaration-de-montreal-pour-un-developpement-responsable-de-lintelligence-artificielle/>