

A photograph of the United Nations Secretariat Building in New York City, viewed through a perspective of many national flags on tall poles. The flags are arranged in two long rows that converge towards the building in the background. The building is a large, light-colored stone structure with a central entrance and a UN emblem on the facade. The sky is a clear, pale blue.

The Geneva Declaration on Artificial Intelligence (A.I.)

Text and copyright © Colum Murphy
The Statecraft Institute, Switzerland 

murphy@globecraft-gsd-ch.com

Document prepared by **Julie de Meyer**,
who contributed to editing, AI-assisted translation, formatting, design, and production.

Globecraft.

Part I – Humanity at a Crossroads

We humans of Earth,
deeply troubled by imminent and mounting threats

- to the survival of the human species, and
- to the survival of our solitary home in the vastness of space;

aware of accelerating dangers now of pandemics, environmental destruction, nuclear war, intercontinental conflict – along with other threats to human destiny as yet unforeseen, are hereby agreed and determined that the advent of powerful A.I. be used for good and not for evil.

1. To this end it is agreed that each significant advance in A.I. be stringently tested for:
 - potential benefits, and
 - potential outcomes harmful to individual liberties.
2. The stringency of A.I. testing shall be no less severe than the level used for the testing of new drugs or important medical procedures.
3. Advanced A.I. testing in advance of acceptable use shall employ a system of licence and imprimatur, so that harmful outcomes be clearly banned and stigmatised ahead of use.
4. It is agreed that A.I. must support United Nations values, democracy, international peace with justice, and the sanctity of the individual – including freedom of thought and expression by that individual, of whatever his or her background, nationality or culture.
5. A committee of no more than 7 persons shall be formed to judge changes in A.I. abilities and potential outcomes for human liberty. Three committee members shall be replaced every 2 years by three new members.
6. It shall be the duty of the A.I. Committee to report regularly, and publicly, on the potential benefits and dangers of A.I.
7. The Committee shall briefly report at least twice a year.
8. The Chairperson of the A.I. Committee shall serve for 2 years, and then be replaced by a new Chairperson from a different continent or regional grouping.
9. A central task and duty of the GAIC (Geneva Artificial Intelligence Committee) shall be to encourage public debate worldwide as to the potential outcomes for humanity of A.I. changes and development.
10. The A.I. Committee shall support beneficial outcomes for humanity: the Committee shall seek and find the proper A.I. place in human destiny and aspirations.

Cruelty shall be the enemy of A.I.

Compassion shall be A.I.'s central and main goal and preoccupation.

Part II – Consciousness

The mission and destiny of humankind is not known to individual humans.

“Why are we here?” is answered by speculation, a mixture of science and faith.

Without certainty,

the role of “consciousness” is central.

Are computers that provide A.I. conscious?

In relation to the above question we must ask what we ourselves mean by “conscious”.

Planet Earth is full of conscious “things”. Animals, sea life, birds and insects are conscious. So also are trees, plants and flowers: every growing thing.

A sunflower, conscious of the sun, turns its head to face and absorb sunlight. A tree may grow moss on only one side of its trunk, protecting it from harsh weather.

Other earthly “things” may be conscious.

So what is our role as humans?

It is agreed that humans must improve their sense of responsibility for all things conscious – humankind thus exerting a Benevolent Dominion over nature, protecting, in a loving or caring manner, all things conscious, to the best of its ability.

A relevant question is: do “things” in nature (whether an animal or a tree) have an awareness of ethics?

A corollary question could then be: is it possible for a human to act unethically while something in nature (say, a tree) acts “correctly”, acting as a tree is supposed to act?

If so, does this then make the tree “superior” to the human? More moral?

If that is so, then from where do we humans derive our “superiority”, the right to exercise our Benevolent Dominion?

The truth is that we humans cannot produce some law or document that “proves” our entitlement to and responsibility for “taking care of nature”.

The best humans can do is to cite the Bible or similar document, from any culture or background, urging humanity to exercise responsibility over all of nature.

In short, any discussion of protecting and nurturing our planetary home inevitably and immediately raises the matters of spirituality and ethics: the key issues of good and evil.

The matter of rightly choosing good over evil cannot be achieved in only a “scientific” manner: choosing good over evil will be a matter for a combination of the spiritual and the scientific, a blending of intuition and information, a seeking of guidance from both brain and heart, a listening to the urgings of our hearts, psyches, souls.

Certain things are truly written on the heart of man.

And the heart of woman.

Consciousness shall be a focus for our search.

It is likely that consciousness is found too in quantum fields.

This virgin field of exploration promises great possibilities of illumination – the imminent discovery of vast new territories and new worlds.

Part III

- **Cruelty is the enemy**
- **Centrally important is *compassion***

It is therefore further decided that:

- The A.I. Committee shall seek, in its rotating membership, persons placing compassion as a priority.
- Compassion shall be a priority – and compassion shall be seen, at the very least, as important as gravity, quantum fields or the speed of light.
- The A.I. Committee shall therefore seek in its membership qualified persons in the fields of spirituality, physics, astronomy, anthropology, philosophy, futurology, medicine, art – and all such fields concerned with the healthy future of humanity, on Earth and in space.
- Common sense shall be important.
- Cruelty is the enemy and a compassionate and successful future for humankind is sought.
- It is agreed that the end does not justify the means: ideologies practising cruelty while claiming to build utopia shall be forbidden, anathema.
- It is therefore hereby also decided, from this city of peace, to advance and propagate the:

GENEVA DECLARATION ON ARTIFICIAL INTELLIGENCE (A.I.)

Copyright Colum Murphy, the Statecraft Institute. Switzerland.

(Improvements/comments are invited from all countries, persons or cultures.)

Partie I – L’humanité à la croisée des chemins

Nous, êtres humains de la Terre,
profondément préoccupés par les menaces imminentes et croissantes

- qui pèsent sur la survie de notre espèce
- et sur celle de notre unique demeure dans l’immensité de l’univers ;

conscients des dangers accélérés que représentent les pandémies, la destruction de l’environnement, la guerre nucléaire, les conflits intercontinentaux – ainsi que d’autres menaces encore imprévisibles pour le destin de l’humanité –,
convenons et affirmons que l’avènement d’une intelligence artificielle puissante doit être orienté vers le bien et non vers le mal.

1. À cette fin, il est décidé que chaque progrès significatif de l’I.A. fasse l’objet d’évaluations rigoureuses portant à la fois sur :
 - ses bénéfices potentiels, et
 - ses conséquences possibles pour les libertés individuelles.
2. Le degré de rigueur exigé pour ces évaluations ne devra pas être inférieur à celui requis pour les essais de nouveaux médicaments ou de procédures médicales majeures.
3. Avant toute utilisation, les systèmes d’I.A. devront être soumis à un régime d’autorisation et d’imprimatur, de manière à interdire et à stigmatiser toute conséquence nocive en amont de leur déploiement.
4. Il est convenu que l’I.A. doit soutenir les valeurs des Nations Unies, la démocratie, la paix internationale avec justice, ainsi que la dignité de la personne humaine – y compris la liberté de pensée et d’expression de chaque individu, indépendamment de son origine, de sa nationalité ou de sa culture.
5. Un comité composé de sept membres au maximum sera institué pour évaluer l’évolution des capacités de l’I.A. et ses répercussions potentielles sur les libertés humaines. Trois membres seront remplacés tous les deux ans par trois nouveaux membres.
6. Le Comité aura pour mission de rendre compte régulièrement, et publiquement, des bénéfices et des dangers potentiels de l’I.A.
7. Il présentera au moins deux rapports succincts par an.
8. La présidence du Comité sera assurée pour une durée de deux ans, avant d’être confiée à une personnalité issue d’un autre continent ou groupe régional.
9. Une mission centrale du GAIC (Geneva Artificial Intelligence Committee) sera de promouvoir un débat public mondial sur les conséquences possibles des évolutions de l’I.A. pour l’avenir de l’humanité.
10. Le Comité soutiendra les développements bénéfiques pour l’humanité et cherchera à définir la juste place de l’I.A. dans le destin et les aspirations humaines.

La cruauté devra être l’ennemi de l’I.A.

La compassion devra constituer son objectif central et primordial.

Partie II – La conscience

La mission et la destinée de l'humanité ne sont pas connues des êtres humains.

À la question : « Pourquoi sommes-nous ici ? », il n'existe que des réponses spéculatives, mêlant science et foi.

En l'absence de certitude,
le rôle de la « conscience » demeure central.

Les ordinateurs dotés d'I.A. sont-ils conscients ?

Cette interrogation nous renvoie à la question préalable : que voulons-nous dire nous-mêmes par « conscience » ?

La planète Terre est peuplée d'êtres conscients : animaux, vie marine, oiseaux, insectes. Mais aussi arbres, plantes et fleurs – toute forme de vie.

Un tournesol, conscient du soleil, tourne sa tête pour capter la lumière. Un arbre peut développer de la mousse d'un seul côté de son tronc pour se protéger des intempéries. D'autres réalités terrestres pourraient être dites conscientes.

Quel est donc notre rôle en tant qu'êtres humains ?

Il est convenu que les humains doivent approfondir leur sens de la responsabilité à l'égard de toutes les formes de conscience – exerçant ainsi une Bienveillance Dominante sur la nature, et protégeant avec soin et respect tout ce qui est conscient, dans la mesure de leurs capacités.

Une question pertinente est de savoir si les « choses » de la nature (un animal ou un arbre, par exemple) possèdent une conscience éthique.

On peut alors se demander : un être humain peut-il agir de façon contraire à l'éthique, tandis qu'un élément de la nature (par exemple, un arbre) agirait « correctement », en conformité avec sa nature propre ?

Si tel est le cas, cela rend-il l'arbre « supérieur » à l'être humain ? Plus moral ?

Et si oui, d'où tirons-nous, nous humains, notre « supériorité » et notre droit à exercer une Bienveillance Dominante ?

La vérité est que nous ne pouvons établir aucune loi ni aucun document qui « prouverait » notre droit et notre responsabilité de « prendre soin de la nature ».

Au mieux, nous pouvons invoquer la Bible ou d'autres textes analogues, issus de diverses cultures et traditions, qui appellent l'humanité à assumer une responsabilité à l'égard de l'ensemble de la nature.

En résumé, toute réflexion sur la protection et le soin de notre demeure planétaire soulève immédiatement les questions de spiritualité et d'éthique : les enjeux fondamentaux du bien et du mal.

Le choix juste du bien contre le mal ne saurait être résolu par la seule science. Ce choix relèvera d'une combinaison entre le spirituel et le scientifique, d'un dialogue entre intuition et information, d'une écoute des messages du cerveau autant que du cœur, d'une ouverture aux appels de nos esprits, de nos psychés, de nos âmes.

Certaines vérités sont inscrites au cœur de l'homme.

Et au cœur de la femme.

La conscience sera un axe central de notre recherche.

Il est probable que la conscience se manifeste également dans les champs quantiques.

Ce domaine vierge d'exploration promet d'immenses découvertes et de nouvelles perspectives d'illumination.

Partie III

- **La cruauté est l'ennemi.**
- **La *compassion* est au centre.**

Il est par conséquent décidé que :

- Le Comité sur l'I.A., dans sa composition tournante, devra s'assurer de la présence de personnes plaçant la compassion au rang de priorité.
- La compassion devra être considérée comme une priorité – au même titre que la gravité, les champs quantiques ou la vitesse de la lumière.
- Le Comité cherchera donc à inclure parmi ses membres des personnalités issues des domaines de la spiritualité, de la physique, de l'astronomie, de l'anthropologie, de la philosophie, de la futurologie, de la médecine, des arts – ainsi que de tous les champs de connaissance relatifs à l'avenir sain et prospère de l'humanité, sur Terre comme dans l'espace.
- Le bon sens devra guider les travaux.
- La cruauté est l'ennemi, et c'est un avenir compatissant et réussi pour l'humanité qui doit être recherché.
- Il est convenu que la fin ne justifie pas les moyens : les idéologies qui recourent à la cruauté tout en prétendant bâtir l'utopie devront être interdites et dénoncées comme telles.
- Il est donc décidé, depuis cette ville de paix, de promouvoir et de diffuser la :

DÉCLARATION DE GENÈVE SUR L'INTELLIGENCE ARTIFICIELLE (I.A.)

© Colum Murphy, The Statecraft Institute. Suisse.

(Avis et contributions sont bienvenus de la part de tous les pays, peuples et cultures.)

Parte I – La humanidad en una encrucijada

Nosotros, los seres humanos de la Tierra,
profundamente preocupados por las amenazas inminentes y crecientes

- que ponen en peligro la supervivencia de nuestra especie
- y de nuestro único hogar en la inmensidad del universo;

conscientes de los peligros que se aceleran, entre ellos las pandemias, la degradación ambiental, la guerra nuclear, los conflictos intercontinentales – así como otras amenazas aún imprevisibles para el destino de la humanidad –,
convenimos y declaramos que el advenimiento de una inteligencia artificial poderosa debe ser encauzado para el bien y nunca para el mal.

1. A tal fin se acuerda que cada avance significativo en la I.A. sea sometido a una evaluación rigurosa que considere:
 - sus beneficios potenciales, y
 - sus posibles repercusiones negativas para las libertades individuales.
2. El rigor de dichas evaluaciones no deberá ser inferior al exigido para la aprobación de nuevos medicamentos o procedimientos médicos de importancia.
3. Antes de su uso, los sistemas de I.A. estarán sujetos a un régimen de autorización y validación que prohíba y estigmatice de manera preventiva toda consecuencia dañina.
4. Se acuerda que la I.A. debe apoyar los valores de las Naciones Unidas, la democracia, la paz internacional con justicia y la dignidad de la persona humana – incluida la libertad de pensamiento y de expresión de cada individuo, cualquiera que sea su origen, nacionalidad o cultura.
5. Se establecerá un comité integrado por un máximo de siete personas para evaluar los cambios en las capacidades de la I.A. y sus posibles efectos en las libertades humanas. Tres de sus miembros serán renovados cada dos años.
6. El Comité tendrá el deber de informar con regularidad, y de manera pública, sobre los beneficios y riesgos de la I.A., presentando al menos dos informes breves al año.
7. La presidencia del Comité será ejercida por un período de dos años y luego pasará a otra persona proveniente de un continente o grupo regional distinto.
8. Una tarea esencial del GAIC (Comité de Ginebra sobre Inteligencia Artificial) será promover un debate público mundial acerca de las posibles consecuencias de la evolución de la I.A. para la humanidad.
9. El Comité respaldará los desarrollos que resulten beneficiosos para la humanidad y buscará determinar el lugar adecuado de la I.A. en el destino y las aspiraciones humanas.

La crueldad deberá ser el enemigo de la I.A.

La compasión deberá constituir su objetivo central y prioritario.

Parte II – La conciencia

La misión y el destino de la humanidad no son conocidos por los propios seres humanos.

A la pregunta: «¿Por qué estamos aquí?», solo se responde con conjeturas, en una mezcla de ciencia y fe.

Ante la ausencia de certezas,
el papel de la «conciencia» resulta central.

¿Son conscientes los ordenadores que sustentan la I.A.?

Ello nos conduce a preguntarnos qué entendemos nosotros mismos por «conciencia».

La Tierra está llena de seres conscientes: animales, vida marina, aves e insectos. También lo son los árboles, las plantas y las flores: toda forma de vida.

Un girasol, consciente del sol, orienta su cabeza para captar la luz. Un árbol puede desarrollar musgo únicamente en un lado de su tronco, protegiéndose así de las inclemencias.

Otras realidades terrestres también podrían considerarse conscientes.

¿Cuál es, entonces, nuestro papel como seres humanos?

Se acuerda que la humanidad debe profundizar su sentido de responsabilidad hacia todas las formas de conciencia – ejerciendo un Dominio Benevolente sobre la naturaleza, protegiendo con cuidado y respeto todo lo consciente, en la medida de sus capacidades.

Una cuestión pertinente es si los «elementos» de la naturaleza (un animal o un árbol, por ejemplo) poseen una percepción ética.

La pregunta correlativa es: ¿puede un ser humano actuar de manera contraria a la ética mientras que un elemento de la naturaleza (digamos, un árbol) actúa «correctamente», de acuerdo con su propia naturaleza?

De ser así, ¿convierte ello al árbol en «superior» al ser humano? ¿Más moral?

Si así fuera, ¿de dónde deriva la humanidad su supuesta «superioridad», su derecho a ejercer ese Dominio Benevolente?

La verdad es que los seres humanos no podemos elaborar ninguna ley ni documento que «pruebe» nuestro derecho y nuestra obligación de «cuidar la naturaleza».

A lo sumo, podemos remitirnos a la Biblia o a otros textos equivalentes de diversas culturas y tradiciones que instan a la humanidad a asumir su responsabilidad sobre la totalidad de la naturaleza.

En suma, cualquier debate sobre la protección y el cuidado de nuestro hogar planetario plantea de inmediato las cuestiones de la espiritualidad y la ética: los problemas fundamentales del bien y del mal.

La elección del bien frente al mal no podrá resolverse únicamente por la ciencia. Será necesario un equilibrio entre lo espiritual y lo científico, una síntesis de intuición e información, una búsqueda de orientación tanto en el cerebro como en el corazón, una escucha atenta de los llamados de nuestro espíritu, de nuestra psique y de nuestra alma.

Hay verdades inscritas en el corazón del hombre.

Y en el corazón de la mujer.

La conciencia será un eje central de nuestra búsqueda.

Es probable que la conciencia se manifieste también en los campos cuánticos.

Ese territorio virgen de exploración promete inmensas posibilidades de descubrimiento e iluminación.

Parte III

- **La crueldad es el enemigo.**
- **La *compasión* es lo esencial.**

En consecuencia, se decide además que:

- El Comité sobre la I.A., en su composición rotativa, deberá incluir a personas que otorguen a la compasión un carácter prioritario.
- La compasión deberá considerarse prioritaria – y equipararse, al menos, a principios fundamentales como la gravedad, los campos cuánticos o la velocidad de la luz.
- El Comité procurará incluir entre sus miembros a personas cualificadas en ámbitos como la espiritualidad, la física, la astronomía, la antropología, la filosofía, la prospectiva, la medicina, el arte – y en todos aquellos campos relacionados con el futuro sano y sostenible de la humanidad, en la Tierra y en el espacio.
- El sentido común deberá ser un principio rector.
- La crueldad es el enemigo, y se aspira a un futuro compasivo y próspero para la humanidad.
- Se acuerda que el fin no justifica los medios: las ideologías que recurran a la crueldad mientras proclaman construir una utopía deberán ser prohibidas y condenadas.
- Se decide, por lo tanto, desde esta ciudad de paz, impulsar y difundir la:

DECLARACIÓN DE GINEBRA SOBRE INTELIGENCIA ARTIFICIAL (I.A.)

© Colum Murphy, The Statecraft Institute. Suiza.

(Se invita a todos los países, pueblos y culturas a formular comentarios y propuestas de mejora.)

第一部分——人类处于十字路口

我们，地球上的人类，
深切关注日益迫近、不断加剧的威胁，
这些威胁危及人类的生存，
并危及我们在浩瀚宇宙中唯一的家园；

认识到当前加速发展的危险，包括流行病、环境破坏、核战争、洲际冲突——以及其他尚不可预见的人类命运威胁，

我们一致同意并郑重声明：强大的人工智能必须用于善，而不得用于恶。

为此决定，每一项重要的人工智能进展必须经过严格测试，测试内容包括：

- 潜在益处；
- 对个人自由可能造成的有害后果。

人工智能的测试严格程度不得低于对新药物或重大医疗程序所要求的水平。

在批准使用前，先进的人工智能必须实行许可和核准制度，从源头禁止并明确标示任何有害结果。

我们一致认为，人工智能必须支持联合国价值观、民主、以公正为基础的国际和平，以及个人的神圣尊严——包括个人的思想和表达自由，无论其出身、国籍或文化背景如何。

将设立一个不超过七人的委员会，评估人工智能能力的变化及其对人类自由的潜在影响。委员会中的三名成员每两年更换一次。

人工智能委员会有义务定期并公开报告人工智能的潜在益处与风险。委员会每年至少简要报告两次。

委员会主席任期两年，届满后由来自其他大洲或区域的成员接任。

日内瓦人工智能委员会（GAIC）的核心职责之一是促进全球公众辩论，讨论人工智能变革与发展对人类可能产生的影响。

委员会应当支持有利于人类的成果，努力探索人工智能在人类命运与理想中的合理定位。

人工智能的敌人应当是残酷。

人工智能的核心目标与首要关切应当是慈悲。

第二部分——意识

人类的使命与命运，并不为个体人类所知。

“我们为何在此？”这一问题只能以推测来回答，融合科学与信仰。

在没有确凿答案的情况下，**“意识”**的作用显得尤为核心。

提供人工智能的计算机是否具有意识？

为回答这一问题，我们必须首先反思自己对“意识”的理解。

地球充满了有意识的“存在”：动物、海洋生物、鸟类和昆虫。树木、植物和花卉亦然——凡是生长的事物。

向日葵因感知太阳而转向以吸收光能；树木可能仅在树干一侧生长苔藓以抵御恶劣天气。

其他自然存在也可能具有意识。

那么，人类的角色是什么？

我们一致同意，人类必须增强对一切有意识存在的责任感——以仁慈的主导姿态守护自然，以关怀和爱护的方式尽力保护一切有意识的存在。

一个相关问题是：自然界中的“事物”（无论是动物还是树木）是否具备伦理意识？

进一步的问题是：是否可能在人类行为违背伦理的同时，自然界中的某个存在（如一棵树）却“正确”地行动，履行它应有的本性？

如果是这样，这是否意味着树木“优于”人类？更具道德？

如果如此，那么人类凭何主张其“优越性”，以及行使仁慈主导权的权利？

事实是，人类无法制定任何法律或文件来“证明”我们有权、有责“照顾自然”。

我们所能做的，最多是援引《圣经》或其他文化传统中的类似典籍，呼吁人类肩负起对自然整体的责任。

简言之，凡是涉及保护与培育我们地球家园的讨论，必然立即触及灵性与伦理问题：善与恶的根本议题。

正确地选择善而非恶，不能仅凭科学来实现。

这一选择需要灵性与科学的结合，需要直觉与信息的融合，需要理智与心灵的共同指引，需要倾听我们内心、精神与灵魂的呼唤。

某些真理确实铭刻在人类的心中。

铭刻在男人的心中。

铭刻在女人的心中。

意识将是我们探索的焦点。

很可能意识也存在于量子场之中。

这一尚未开发的领域蕴藏着广阔的启示前景——人类即将发现广袤的新天地和新世界。

第三部分

残酷是敌人。

慈悲是核心。

因此，进一步决定：

人工智能委员会在轮换任命时，应确保吸纳将慈悲视为优先事项的人士。

慈悲应被视为首要原则——其重要性至少应与重力、量子场或光速相提并论。

委员会成员应当包括来自灵性、物理学、天文学、人类学、哲学、未来学、医学、艺术等领域的专家，以及所有与人类在地球与宇宙中的健康未来相关的领域代表。

常识应当成为重要的指导原则。

残酷是敌人。人类应当追求一个以慈悲为核心的成功未来。

我们一致同意，目的不能证明手段的正当性：任何借助残酷手段、却声称建设乌托邦的意识形态，必须被禁止并予以谴责。

因此，从这座和平之城郑重宣布并推广：

《日内瓦人工智能宣言》

© Colum Murphy, 国家治理研究所。瑞士。

（欢迎各国、各民族、各文化提出改进意见与建议。）

الجزء الأول - الإنسانية عند مفترق طرق

نحن، بشر الأرض،

المضطربون بعمق إزاء التهديدات الوشيكة والمتزايدة

لبقاء الجنس البشري،

ولبقاء موطننا الوحيد في اتساع الفضاء؛

وإدراكاً للأخطار المتسارعة المتمثلة في الجوائح وتدمير البيئة والحرب النووية والنزاعات بين القارات - فضلاً عن غيرها من التهديدات غير

المتوقعة لمصير الإنسانية،

نتفق ونقر بأن ظهور الذكاء الاصطناعي المتقدم يجب أن يُسَخَّر للخير لا للشر.

ولهذا الغرض، يُتفق على أن يخضع كل تقدّم جوهري في مجال الذكاء الاصطناعي لاختبارات صارمة تشمل:

- المنافع المحتملة، و
- النتائج الضارة المحتملة على الحريات الفردية.

ولا يجوز أن تقل صرامة اختبارات الذكاء الاصطناعي عن المستوى المعتمد لاختبار الأدوية الجديدة أو الإجراءات الطبية الهامة.

وقبل السماح بالاستخدام، يتعين أن يخضع الذكاء الاصطناعي المتقدم لنظام ترخيص وموافقة، بحيث تُحظر النتائج الضارة ويجري

التنبيه إليها قبل وقوعها.

ويُتفق على أن يدعم الذكاء الاصطناعي قيم الأمم المتحدة والديمقراطية والسلام الدولي القائم على العدل، وكرامة الفرد - بما في ذلك

حرية الفكر والتعبير لكل إنسان، أيّاً كان أصله أو جنسيته أو ثقافته.

ويُنشأ لجنة لا يزيد عدد أعضائها على سبعة أشخاص لتقييم التغيرات في قدرات الذكاء الاصطناعي وما قد يترتب عليها من آثار على

حرية الإنسان. ويُستبدل ثلاثة أعضاء كل سنتين بثلاثة أعضاء جدد.

وتُنَاط باللجنة واجبات منها رفع تقارير منتظمة وعلمية بشأن المنافع والمخاطر المحتملة للذكاء الاصطناعي. ويُقدّم تقريران موجزان

على الأقل سنوياً.

ويتولى رئاسة اللجنة رئيس لمدة سنتين، ثم يُستبدل برئيس جديد من قارة أو مجموعة إقليمية أخرى.

وتتمثل مهمة محورية للجنة جنيف للذكاء الاصطناعي (GAIC) في تشجيع النقاش العام على نطاق العالم حول النتائج المحتملة

لتطورات الذكاء الاصطناعي بالنسبة لمستقبل الإنسانية.

وتعمل اللجنة على دعم النتائج المفيدة للبشرية، والبحث عن الموقع المناسب للذكاء الاصطناعي في مصير الإنسان وتطلعاته.

ينبغي أن تكون القسوة عدو الذكاء الاصطناعي.

وينبغي أن تكون الرحمة هدفه المركزي والغاية الأولى له.

الجزء الثاني - الوعي

إن رسالة الإنسان ومصيره غير معروفين للبشر الأفراد.

ويظل سؤال: "لماذا نحن هنا؟" سؤالاً يُجاب عليه بالتخمين، بمزيج من العلم والإيمان.

وبغياب اليقين، يبقى دور "الوعي" محورياً.

هل الحواسيب التي توّجّر الذكاء الاصطناعي واعية؟

وعند طرح هذا السؤال، علينا أن نسأل أنفسنا: ما الذي نعتيه نحن أنفسنا بلفظ "الوعي"؟

إن كوكب الأرض مليء بالكائنات الواعية: الحيوانات والكائنات البحرية والطيور والحشرات. وكذلك الأشجار والنباتات والأزهار - كل ما ينمو. فزهرة دوّار الشمس، بوعيتها بالشمس، تدير وجهها لتمتص ضوءها. والشجرة قد ينمو الطحلب على جانب واحد من جذعها لتحمي نفسها من قسوة الطقس.

وقد تكون هناك "أشياء" أخرى في الطبيعة واعية كذلك.

فما هو إذن دورنا كبشر؟

يُتفق على أن يتعين على الإنسان أن يعزز إحساسه بالمسؤولية إزاء كل ما هو واعٍ - وأن يمارس سيادة رحمة على الطبيعة، فيحمي جميع الكائنات الواعية بحب وعناية، قدر استطاعته.

والسؤال ذو الصلة هو: هل للكائنات في الطبيعة (سواء حيوان أو شجرة) إدراك للأخلاق؟

ويترتب على ذلك سؤال آخر: هل يمكن أن يتصرف الإنسان تصرفاً منافياً للأخلاق، بينما يتصرف شيء في الطبيعة (كالشجرة مثلاً) تصرفاً "صحيحاً"، أي وفق ما خلّق ليكون عليه؟

وإذا كان ذلك ممكناً، فهل يجعل هذا الشجرة "أسمى" من الإنسان؟ أكثر أخلاقية؟ وإذا كان الأمر كذلك، فمن أين يستمد الإنسان "تفوقه" وحقه في ممارسة السيادة الرحمة؟

والحقيقة أن الإنسان لا يستطيع أن يصدر قانوناً أو وثيقة "تثبت" أحقيته في رعاية الطبيعة. وأقصى ما يمكن أن يفعله هو أن يستشهد بالكتاب المقدس أو بنصوص مماثلة من أي ثقافة أو خلفية، تحت الإنسانية على تحمل المسؤولية تجاه الطبيعة كلها.

وخلاصة القول: إن أي نقاش حول حماية بيتنا الكوكبي ورعايته يثير فوراً مسألتَي الروحانية والأخلاق: قضيتي الخير والشر.

إن الاختيار الصحيح للخير على الشر لا يمكن أن يُحسم بالعلم وحده. بل سيكون مسألة تمزج بين الروحاني والعلمي، بين الحدس والمعرفة، بين إرشاد العقل وإلهام القلب، بين الاستماع إلى نداء أرواحنا ونفوسنا وضمائرنا.

إن أشياء معينة منقوشة في قلب الرجل.

وفي قلب المرأة.

وسيبقى الوعي محور بحثنا.

ومن المرجح أن يوجد الوعي كذلك في الحقول الكمومية.

وهذا المجال البكر من الاستكشاف يبشر بإمكانات هائلة من الإشراف واكتشاف عوالم جديدة وآفاق رحبة.

الجزء الثالث

القسوة هي العدو.

والرحمة هي الجواهر.

وبناءً عليه، يُقرر كذلك ما يلي:

ينبغي للجنة الذكاء الاصطناعي، في إطار عضويتها الدورية، أن تضم أشخاصاً يعتبرون الرحمة أولوية.

ويجب اعتبار الرحمة أولوية - على الأقل مساوية للجاذبية والحقول الكمومية وسرعة الضوء.

وعليه، تسعى اللجنة إلى أن تضم بين أعضائها أشخاصاً مؤهلين في مجالات الروحانية، الفيزياء، الفلك، الأنثروبولوجيا، الفلسفة، استشراف المستقبل، الطب، الفنون - وجميع المجالات ذات الصلة بمستقبل البشرية السليم على الأرض وفي الفضاء.

وينبغي أن يكون المنطق السليم ركناً أساسياً.

القسوة هي العدو، والمستقبل المنشود للبشرية هو مستقبل يقوم على الرحمة والنجاح.

ويُتفق على أن الغاية لا تبرر الوسيلة: فأَيُّ أيديولوجيات تمارس القسوة بزعم بناء المدينة الفاضلة يجب أن تُحظر وتُبدن.

ومن ثم، ومن هذه المدينة التي ترمز إلى السلام، يُقرر المضي في إعلان ونشر:

إعلان جنيف بشأن الذكاء الاصطناعي

© كولوم مورفي، معهد فن الحكم. سويسرا.

(تُرَجَّب الملاحظات والمقترحات من جميع البلدان والأشخاص والثقافات).

Часть I – Человечество на перепутье

Мы, люди Земли,
глубоко обеспокоенные надвигающимися и возрастающими угрозами
выживанию человеческого рода и
выживанию нашего единственного дома в безграничности космоса;

осознавая ускоряющиеся опасности пандемий, разрушения окружающей среды, ядерной войны,
межконтинентальных конфликтов, а также иных, пока непредвиденных угроз судьбе человечества,
соглашаемся и решительно заявляем, что появление мощного искусственного интеллекта должно
служить во благо, а не во зло.

В этой связи согласовано, что каждый значительный прогресс в области искусственного интеллекта
должен подвергаться строгому тестированию на:

- потенциальную пользу, и
- возможные последствия, вредные для индивидуальных свобод.

Строгость тестирования ИИ должна быть не ниже уровня, применяемого при испытаниях новых
лекарственных средств или важных медицинских процедур.

Использование передовых систем ИИ должно предваряться лицензированием и официальным допуском,
чтобы вредные последствия были заранее запрещены и признаны недопустимыми.

Согласовано, что искусственный интеллект обязан поддерживать ценности Организации Объединённых
Наций, демократию, международный мир на основе справедливости и неприкосновенность личности –
включая свободу мысли и выражения каждым человеком, независимо от его происхождения,
национальности или культуры.

Создаётся комитет численностью не более семи человек для оценки изменений в возможностях ИИ и их
последствий для человеческих свобод. Трое членов комитета заменяются каждые два года.

Обязанностью Комитета по ИИ является регулярное и публичное информирование о потенциальных
преимуществах и опасностях ИИ.

Краткий доклад представляется не реже двух раз в год.

Председатель Комитета по ИИ исполняет свои обязанности в течение двух лет, после чего заменяется
представителем с другого континента или из иного регионального объединения.

Центральной задачей и обязанностью ГАИК (Женевского комитета по искусственному интеллекту)
является стимулирование всемирных общественных дебатов о возможных последствиях развития ИИ для
человечества.

Комитет по ИИ поддерживает благоприятные для человечества результаты и стремится определить
надлежащее место ИИ в человеческой судьбе и стремлениях.

Жестокость должна быть врагом ИИ.

Сострадание должно стать его центральной целью и задачей.

Часть II – Сознание

Миссия и предназначение человечества не известны отдельным людям.

На вопрос: «Зачем мы здесь?» – даются лишь предположительные ответы, сочетающие науку и веру. При отсутствии определённости роль «сознания» является ключевой.

Являются ли компьютеры, обеспечивающие работу ИИ, сознательными?

Чтобы ответить на этот вопрос, мы должны определить, что сами понимаем под «сознанием».

Планета Земля полна сознательных «сущностей». Животные, морская жизнь, птицы и насекомые обладают сознанием. Так же, как деревья, растения и цветы – всё живое.

Подсолнух, осознавая солнце, поворачивает голову, чтобы впитать свет. Дерево может покрыться мхом только с одной стороны, защищая себя от суровой погоды.

Иные земные существа также могут обладать сознанием.

Какова же наша роль как людей?

Согласовано, что человечество должно повышать чувство ответственности за все сознательные сущности, проявляя Доброе Попечение над природой, защищая всё живое с любовью и заботой в пределах своих возможностей.

Возникает вопрос: обладают ли «существа» природы (животные, деревья) осознанием этики?

Сопутствующий вопрос: возможно ли, что человек поступает неэтично, в то время как некий природный объект (например, дерево) действует «правильно», в соответствии со своей природой?

Если да, делает ли это дерево «выше» человека, более нравственным?

Если это так, то откуда человек черпает своё «превосходство» и право на Доброе Попечение?

Истина заключается в том, что человечество не может создать закон или документ, который «доказывал» бы его право и обязанность «заботиться о природе».

Лучшее, что люди могут сделать, – это обратиться к Библии или аналогичным текстам любой культуры, которые призывают человечество к ответственности за всю природу.

Иными словами, всякая дискуссия о защите и сохранении нашей планеты неизбежно и сразу затрагивает вопросы духовности и этики – ключевые вопросы добра и зла.

Выбор в пользу добра не может быть достигнут исключительно «научным» путём: он требует сочетания духовного и научного начал, соединения интуиции и информации, поиска руководства как разумом, так и сердцем, прислушивания к внутренним устремлениям души.

Определённые истины действительно написаны на сердце человека.
И на сердце женщины.

Сознание должно стать предметом нашего поиска.

Вероятно, сознание также связано с квантовыми полями.

Это новое направление исследований сулит великие возможности озарения – открытие обширных новых пространств и миров.

Часть III

Жестокость – враг.

Сострадание – центральная ценность.

В этой связи далее постановляется:

Комитет по ИИ должен обеспечивать, чтобы в его ротационном составе участвовали лица, для которых сострадание является приоритетом.

Сострадание должно быть признано приоритетом – и рассматриваться, по меньшей мере, столь же важным, как гравитация, квантовые поля или скорость света.

Таким образом, Комитет по ИИ должен включать специалистов в областях духовности, физики, астрономии, антропологии, философии, футурологии, медицины, искусства – и всех иных областях, связанных со здоровым будущим человечества на Земле и в космосе.

Здравый смысл должен иметь важное значение.

Жестокость – враг, и человечество стремится к сострадательному и успешному будущему.

Согласовано, что цель не оправдывает средства: идеологии, практикующие жестокость под предлогом построения утопии, должны быть запрещены и признаны недопустимыми.

Настоящим также постановляется, из этого города мира, продвигать и утверждать:

ЖЕНЕВСКАЯ ДЕКЛАРАЦИЯ ОБ ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ (ИИ)

Copyright Colum Murphy, The Statecraft Institute, Швейцария.

(Предложения и замечания приветствуются от всех стран, лиц и культур.)



Text and copyright © Colum Murphy
The Statecraft Institute, Switzerland 

murphy@globecraft-gsd-ch.com

*Document prepared by **Julie de Meyer**,
who contributed to editing, AI-assisted translation, formatting, design, and production.*

Globecraft.