



Declaración final sobre Robótica, Inteligencia Artificial y Humanidad

Ciencia, Ética y Políticas



PROBLEMÁTICAS Y AGENDA

1. Los avances en la robótica y el aprendizaje de máquinas (a menudo llamado «inteligencia artificial, IA») están suscitándose a una velocidad cada vez mayor. Su impacto en el funcionamiento de las sociedades y las economías es significativo, y sus beneficios y desventajas para la humanidad son motivo de mucho debate. Este pujante campo de la ciencia y la tecnología exige nuestra cuidadosa atención. Las Academias Pontificias, por lo tanto, han organizado esta conferencia,¹ la cual ha reunido a colegas provenientes de diversas disciplinas.
2. La inteligencia artificial y las aplicaciones de las tecnologías robóticas habrán de tener implicancias trascendentes para las economías, el tejido social y la cultura. Además de analizar las actuales fronteras de la investigación en la materia, hemos revisado sus potenciales impactos en el bienestar social, y sus implicancias en términos de ética y de políticas.
3. La IA y la robótica son muy prometedoras a la hora de abordar algunos de nuestros más complejos problemas sociales, económicos y ambientales, como por ejemplo el cambio climático y el logro de los Objetivos de Desarrollo Sostenible de la ONU (ODS) llegado el año 2030.
4. También hay tendencias a largo plazo en la materia, con consecuencias que a la larga pueden poner en cuestión el lugar del ser humano en la sociedad.

5. Revisten creciente preocupación los riesgos para la paz planteados por las nuevas tecnologías bélicas (p. ej., ataques cibernéticos y armas autónomas), situación que amerita implantar nuevas normas internacionales en materia de seguridad.
6. Es fundamental aclarar los aspectos éticos y religiosos de la IA y la robótica a la hora de orientar las políticas regulatorias que puedan hacer falta en materia de aplicaciones y evolución de estas tecnologías de cara al futuro.

PROBLEMÁTICAS FUNDACIONALES EN IA Y ROBÓTICA

7. El campo de la inteligencia artificial ha desarrollado una rica variedad de **enfoques y marcos teóricos** por un lado, y una serie de **aplicaciones prácticas** cada vez más sorprendentes, por otro. La IA tiene el potencial de aportar avances en todas las áreas de la ciencia y la sociedad. Puede ayudarnos a superar algunas de nuestras limitaciones cognitivas, y a resolver problemas complejos. Mientras que los datos en cantidades gigantescas representan un desafío para las habilidades cognitivas del ser humano, los *big data*, o macrodatos, ofrecen oportunidades inauditas para las ciencias y las humanidades. El potencial traslacional de los *big data* es enorme, por ejemplo en los campos de la medicina, la salud pública, la educación, y la gestión de sistemas complejos en general (caso de la biosfera, la geosfera, y la economía). Sin embargo, el quehacer científico basado en estos macrodatos sigue siendo puramente empírico, lo que plantea el interrogante de cuáles son los mecanismos causales subyacentes que son responsables de generar patrones.
8. En combinación con las **interfaces cerebro-computadora y la robótica**, la IA ya es de gran ayuda para los pacientes con trastornos sensoriales o motores, y facilita el cuidado de las personas con discapacidad. Al brindar herramientas novedosas para la adquisición del conocimiento, la IA es capaz de aportar cambios radicales en la educación, y puede facilitar el acceso al saber. Puede también haber sinergias que resulten de la interacción robot-robot, al igual que posibles sinergias positivas donde humanos y robots trabajen codo a codo en determinadas tareas.
9. Entre las problemáticas fundacionales de la IA y la robótica, está la cuestión de si las máquinas pueden, hipotéticamente, adquirir **capacidades tales como la conciencia**. Actualmente este interrogante se debate desde las perspectivas de las ciencias naturales, la teoría social y la filosofía, con posturas divergentes; sigue siendo una pregunta sin respuesta, en gran medida porque hay una diversidad de definiciones del concepto del «conciencia». También hay quienes se han preguntado si acaso el énfasis en las capacidades computacionales y de compilación de la IA, catalogadas como sobrehumanas, no enmascaran las muchas limitaciones de los sistemas artificiales.
10. Son muchos los que sostienen que los **robots no pueden ser considerados personas**, de

modo que no deberían ser agentes poseedores de libre albedrío, ni tener derechos. Sin embargo, hay quienes argumentan que los conceptos de «mando y control» quizás no sean adecuados para la relación humano-robot, y otros incluso se preguntan si acaso no podría considerarse el concepto de «ciudadanía electrónica». **La filosofía y la teología cristianas** establecen que el alma humana es intrínsecamente incorruptible. Esta es la base metafísica según la cual la persona humana es en sí misma libre y capaz de un orden ético, y emana de las fuerzas de la naturaleza. Como sujeto espiritual, el ser humano es *imago Dei*. Bajo la égida de la filosofía cristiana, la IA y los robots no pueden ser considerados personas, con lo cual nunca habrán de poseer (ni deberían) libertad humana ni un alma espiritual, y no pueden ser vistos como «imágenes de Dios». Quizás sí se los pueda tomar como «imágenes del ser humano», pudiendo ser creados por el hombre como instrumentos para el bien de la sociedad.

11. Una de las líneas de investigación sobre aprendizaje de máquinas apunta a identificar las justificaciones fundacionales del **diseño de los agentes cognitivos**; tales justificaciones permitirían derivar teoremas que definan las posibilidades y las limitaciones de los agentes inteligentes. Los **agentes cognitivos** actúan en un entorno abierto y parcial o completamente desconocido, con el fin de cumplir objetivos. Algunos conceptos clave del marco fundacional de la IA son los siguientes: agentes, entornos, recompensas, puntajes locales, puntajes globales, el modelo exacto de interacción entre agentes y entornos, y la especificación de los recursos computacionales disponibles de los agentes y sus entornos. Un «agente inteligente» puede definirse como aquel que cumple objetivos en una amplia gama de entornos. El tema de la cognición con recursos limitados ameritaría un mayor análisis si la idea es llegar a un marco fundacional para la inteligencia artificial que sea sólido e integral.

12. Un aspecto cardinal del aprendizaje a partir de la experiencia es la representación y el procesamiento del conocimiento incierto. Ante la falta de supuestos deterministas sobre el mundo, no es posible derivar ninguna conclusión lógica no trivial acerca del pasado o del futuro. En consecuencia, reviste interés fundacional analizar la **estructura de la incertidumbre** como cuestión en sí misma. Todavía falta mucho en materia de investigación para poder trasladar las propuestas preliminares y los métodos formales del reino teórico al ámbito de la ingeniería de las soluciones algorítmicas eficientes.

13. Algunos resultados recientes señalan una estrecha **conexión entre la capacidad de aprendizaje y la demostrabilidad**, lo cual reduce la cuestión de lo que efectivamente puede aprenderse a los interrogantes fundacionales que propone la matemática en torno a los axiomas de existencia de conjuntos. Los resultados de la matemática inversa —rama de la lógica matemática que se encarga de analizar teoremas en referencia a los axiomas de existencia de conjuntos necesarios para demostrarlos— pueden servir para ilustrar las implicancias de los marcos planteados por el aprendizaje de máquinas. En general, las técnicas de verificación de modelos y de pruebas se vuelven más importantes en la medida que aumenta la importancia de las tareas confiadas a agentes inteligentes.

14. Hasta hace poco, las ciencias matemáticas básicas tenían, con suerte, unos pocos interrogantes éticos a resolver. Ahora, por el contrario, dado que son centrales al desarrollo de la IA, es fundamental que los matemáticos y los diseñadores de *software* consideren las implicancias éticas de su quehacer. A la luz de las cuestiones que cada vez más se plantean en torno a la confiabilidad de los sistemas autónomos, **los responsables de desarrollar soluciones de IA tienen la responsabilidad —la cual idealmente debería convertirse en una obligación legal— de crear sistemas robóticos confiables y controlables.**

CIENCIA, INGENIERÍA E INTERACCIÓN HUMANO- IA/ROBOT

15. Se están llevando a cabo investigaciones de gran porte en las áreas que nos definen como humanos: tal es el caso del lenguaje, el procesamiento simbólico, el aprendizaje en un intento (*one-shot learning*), la auto-evaluación, los juicios de confianza, la programación inductiva, la concepción de metas, y la integración de módulos existentes en una arquitectura inteligente dominante y con múltiples propósitos. Los agentes computacionales entrenados mediante el aprendizaje por refuerzo y el aprendizaje profundo demuestran una capacidad de desempeño notable en tareas que hasta hace poco se consideraban inviables para ellos. Si bien todavía no hay bases sólidas para una teoría general de los agentes cognitivos computacionales, el progreso conceptual y práctico de la inteligencia artificial ha sido tal que **los interrogantes éticos y en materia de seguridad humana, y el impacto de estas tecnologías en la sociedad en su conjunto, se vuelven problemáticas urgentes.** Un ejemplo de ello son los sistemas de IA que deducen los sentimientos de las personas a partir de los datos de reconocimiento facial.

16. La difusión de la robótica modifica profundamente los vínculos humanos y sociales en todos los aspectos de la sociedad, desde la familia hasta el ámbito laboral y la esfera pública. Estas modificaciones conllevan una hibridación entre las características inherentemente humanas de estos vínculos y sus características artificiales, o sea, entre la realidad analógica y la virtual. Por consiguiente, es necesario acentuar la investigación científica en torno a los efectos en la sociedad de delegar aspectos relevantes de la organización social a los robots y la IA. Un objetivo de dicha investigación debería consistir en entender cómo es posible gobernar los procesos de cambio más importantes y producir los bienes relacionales que concretan una virtuosa plenitud humana en el marco de un desarrollo social sostenible y justo.

17. Observamos que **el progreso acelerado de la ingeniería robótica** está transformando algunas industrias de cabo a rabo (la así llamada Industria 4.0). La evolución del Internet de las cosas (IdC), concepto según el cual las máquinas se comunican entre sí y aprenden unas de otras, ofrece avances de envergadura para servicios tales como los bancarios y los financieros. Las interacciones robot-robot y robot-humano se están volviendo cada vez más intensivas. Y sin embargo, los sistemas de IA son difíciles de poner a prueba y de validar. Esto plantea desafíos a la hora de confiar en estas tecnologías. Están surgiendo **problemas en torno a cómo legislar,** establecer la propiedad de los datos, asignar responsabilidades y asegurar la transparencia de los

algoritmos, y todo esto exige arreglos institucionales legítimos.

18. Se puede distinguir entre robots mecánicos, diseñados para realizar tareas rutinarias en el ámbito industrial, y la capacidad de la IA de asistir en programas de bienestar social, procedimientos médicos, sistemas de movilidad con uso eficiente de la energía, tareas educativas e investigación científica. Mientras los asistentes inteligentes pueden beneficiar a niños y adultos por igual, también conllevan riesgos, porque se desconoce su impacto en el cerebro cuando este está en proceso de desarrollo, y porque **las personas pueden llegar a perder la motivación en las áreas donde la IA aparente ser superior.**

19. En las últimas décadas, el campo de la robótica ha dado origen a **un sinnúmero de aplicaciones al servicio del ser humano.** Un ejemplo paradigmático, entre tantos otros, es el de los brazos robóticos, que deben responder a desafíos tales como la precisión, la sensibilidad, la robustez y la seguridad en el agarre. Aunque los brazos robóticos aún están muy lejos de sus equivalentes humanos, su desempeño ha mejorado mucho gracias a los nuevos métodos de control. Además están surgiendo aplicaciones prometedoras de los sistemas de manipulación remota en áreas tan variadas como la medicina, la producción industrial y la movilidad.

20. **La IA puede servir para optimizar el buen gobierno.** Algunos ejemplos son su aplicación en la identificación y prevención de transacciones ilegales, como es el caso de los fondos emanados de delitos tales como el narcotráfico, la trata o el tráfico de órganos. Sin embargo, cuando la IA está en manos de empresas privadas exclusivamente, se corre el riesgo de que sus réditos no se redistribuyan de forma equitativa. Estas nuevas tecnologías, entonces, no deben transformarse en instrumentos para esclavizar a las personas o marginar aún más a los pobres.

LA ROBÓTICA Y SU POTENCIAL DE MODIFICAR EL TRABAJO, LA AGRICULTURA, LA POBREZA, LA DESIGUALDAD Y LA ECOLOGÍA

21. Hemos revisado las aplicaciones de la IA (y de las tecnologías que de ella surjan) en la medicina y la atención de la salud, en la movilidad y el transporte, y en la industria y la agricultura. Se detectaron grandes oportunidades, y se prestó mucha atención a las aplicaciones de la IA y la robótica en cada uno de estos ámbitos del quehacer humano por separado. Sin embargo, las perspectivas sectoriales en materia de IA y de robótica tienen sus limitaciones. Pareciera necesario contar con un panorama más abarcador de las **relaciones entre las aplicaciones** y las políticas públicas que facilitan **la justicia y la equidad** a lo largo de todos los aspectos de estas novedosas tecnologías.

22. Salvo que se canalice en pro del bien común, la IA plantea grandes inquietudes para la economía y la estabilidad de la sociedad. Varios **puestos de trabajo pueden desaparecer** a manos de los dispositivos computarizados, con la resultante exacerbación de las brechas de conocimiento y la disparidad de los ingresos entre los diferentes sectores de la sociedad. Los

avances en la automatización y la creciente disponibilidad de mano de obra artificial, particularmente en los sectores agrícola e industrial, pueden socavar las posibilidades de empleo en las economías emergentes. Como consecuencia de las vinculaciones dentro de las cadenas mundiales de valor, los trabajadores de los países de bajos ingresos también pueden verse afectados por los robots introducidos en las naciones de altos ingresos, situación que podría reducir la necesidad de tercerizar trabajos rutinarios en regiones antaño caracterizadas por sus bajos salarios. No obstante, el uso de los robots también podría aumentar la demanda de mano de obra mediante la reducción de los costos de producción, algo que podría conducir a la expansión industrial. Es extremadamente necesario contar con estimaciones confiables de los nuevos puestos de trabajo que se crearían en las industrias del diseño y la fabricación de robots.

23. Las implicancias de la robótica en materia de trabajo y empleo constituyen todo un tema cuando de políticas públicas se trata. Estas deberían apuntar a **garantizar las medidas necesarias en términos de seguridad social** para los trabajadores afectados, y al mismo tiempo, a invertir en el desarrollo de las habilidades necesarias para sacar ventaja de los nuevos empleos que se generen. El Estado debería ser capaz de redistribuir las ganancias percibidas por el trabajo realizado por los robots. Tal redistribución de fondos, por ejemplo, podría servir para volver a capacitar a los individuos afectados, de modo que puedan seguir integrando la población activa. En este sentido, cabe recordar que muchas de estas innovaciones tecnológicas se concretaron con el apoyo de fondos públicos.

24. **Los robots, la IA y el capital digital en general podrían ser considerados como base imponible.** Actualmente esto no sucede: la mano de obra humana está gravada, no así la robótica. De este modo, los sistemas robóticos gozan de un subsidio indirecto, pues al no ser objeto de impuestos, sirven para que las empresas reduzcan su carga fiscal. Estas distorsiones, que desfavorecen al trabajador humano y favorecen la inversión en soluciones robotizadas, deberían revertirse.

25. Hemos observado que **hace falta poner más atención** en las implicancias, las oportunidades y los riesgos que **la IA y la robótica plantean para el desarrollo sostenible y los pobres**. En particular, cuando se trata de países de bajos ingresos, de personas marginadas, y de mujeres, es necesario considerar a fondo las implicancias de estas tecnologías en los programas y las políticas de Estado. Los recursos pedagógicos que ofrece la IA revisten una oportunidad interesante para muchas regiones de bajos ingresos. Una gran proporción de la población más pobre del planeta vive en pequeñas viviendas rurales, sobre todo en África y en el sur y el este de Asia, y para estas personas hay una gran diferencia entre tener y no tener acceso a las principales tecnologías digitales y a la inteligencia artificial. Algunos ejemplos, entre muchos, son la obtención de los certificados de propiedad de tierras a través de la tecnología *blockchain*, y el uso de tecnologías de precisión en manejo de suelos y cultivos.

26. Debería prestarse más atención a los **impactos ambientales** tanto directos como indirectos. El

monitoreo de los sistemas terrestres y acuáticos a través de sensores remotos inteligentes puede optimizarse muchísimo, y servir para evaluar los cambios en materia de biodiversidad y los impactos de la intervención humana en los ecosistemas. También está la problemática de la contaminación por desechos electrónicos, que los países industrializados descartan en los territorios de sus pares de bajos ingresos. Esta situación exige una atención urgente, al igual que la huella de carbono que dejan la IA y la robótica.

ROBÓTICA, INTELIGENCIA ARTIFICIAL Y CONFLICTOS ARMADOS

27. En el ámbito de los conflictos armados, los sistemas basados en la IA (incluidos los robots) pueden servir para una variedad de fines, entre otros, rescatar efectivos malheridos, monitorear el cumplimiento de las leyes de guerra y las reglas de empuñamiento, mejorar el conocimiento del estado de situación y la planificación de contiendas bélicas, y elegir blancos de ataque. Mientras la última categoría es la que plantea **los interrogantes morales más inquietantes**, en todos los casos, y antes de implementar estas tecnologías en entornos bélicos, deben estudiarse con cuidado los riesgos que ellas encierran, además de las consecuencias que podría tener la remoción de algunas barreras contra la guerra.

28. El miedo a quedar rezagados en el afán de desarrollar nuevas aplicaciones militares no debe ser excusa para soslayar las investigaciones en materia de seguridad, las puestas a prueba de las tecnologías, y la debida capacitación del personal. Dado que el diseño armamentístico, dejando atrás las infraestructuras de gran escala, está apostando a sistemas autónomos, descentralizados y en miniatura, los riesgos de los efectos destructivos por causa de defectos de diseño y posibles fallas aumentan en forma exponencial comparados con la mayoría de los sistemas que están actualmente operativos. Este mayor potencial de externalidades negativas debe compensarse con una proporcional inversión en seguridad física y capacitación. La IA debe mejorar, no socavar, el ejercicio del buen y moral criterio por parte del personal militar.

29. Debe invertirse un **esfuerzo internacional** mancomunado en pro de identificar las aplicaciones específicas de la IA que puedan conducir a la exacerbación de los conflictos armados. Los Estados deberían acordar medidas concretas para reducir el riesgo de las escaladas bélicas facilitadas por estas tecnologías. Debe prestarse atención no solo a los peligros de la IA como reemplazo del ser humano en el ámbito militar, sino también a los riesgos que puede revestir la misma con respecto a la «reflexión estratégica» en el seno de los conflictos.

30. Respecto de los sistemas de armas autónomas, dadas las competencias técnicas actuales (y las que habrá en un futuro próximo), no debería utilizarse ningún sistema que funcione sin supervisión. **Deben mantenerse las líneas de mando donde el componente humano se haga responsable**, de modo de garantizar el cumplimiento de las leyes de guerra reconocidas a nivel internacional, y de manera de sancionar las transgresiones.

DIMENSIONES SOCIALES, ÉTICAS, RELIGIOSAS Y REGULATORIAS DE LA INTELIGENCIA ARTIFICIAL Y LA ROBÓTICA

Además de las problemáticas ya señaladas, debería prestarse atención a las siguientes:

31. Todo desarrollo de **máquinas inteligentes financiado con fondos públicos debe estar orientado al bien común**. El impacto en los bienes y los servicios públicos, así como en la salud, la educación, la felicidad y la sostenibilidad, debe tener máxima prioridad. La IA puede generar sesgos inesperados o consecuencias inhumanas —incluyendo la segmentación de la sociedad y los sesgos raciales y de género—, que es necesario abordar antes de que ocurran. Estas problemáticas pueden ser de índole nacional o mundial: en el último caso, es necesaria una mayor atención de parte de la ONU.

32. En materia de privacidad, acceso a nuevos conocimientos y derecho a la información, **los pobres son particularmente vulnerables** debido a su falta de poder y de voz. La IA y la robótica deben ir de la mano del empoderamiento de los pobres a través de la información, la educación y la inversión en las habilidades que sea necesario fortalecer.

33. **La temática del trabajo** es central a la Doctrina Social de la Iglesia. Como reza *Laborem exercens*, tener un trabajo pleno de sentido es algo inherente a la dignidad humana. Es inaceptable que la irrupción de las máquinas eche por tierra este concepto cardinal, pero por otro lado, la reducción por parte de la IA de la carga de trabajo y de los riesgos para la salud en materia ocupacional representa un avance constructivo.

34. Las políticas de los Estados deberían apuntar a **compartir los beneficios de la productividad** mediante una redistribución de las ganancias que, en lugar de subsidiar el uso de robots, considerase la aplicación de impuestos sobre el capital (digital), y la reducción de las horas de trabajo dedicadas a tareas rutinarias.

35. Merecen atención los riesgos del **uso de la IA para manipular** la opinión pública e interferir en las elecciones gubernamentales. A este fin, hacen falta mecanismos de control a nivel nacional e internacional.

36. La inteligencia artificial y la robótica ofrecen grandes oportunidades al tiempo que encierran riesgos; por consiguiente, las instituciones públicas pertinentes **deben diseñar normativas que sean adecuadas**, que no impliquen trabas a las oportunidades, que no estimulen el sesgo y que no inciten a tomar riesgos excesivos. Para esto hace falta un marco en el que el discurso social público e inclusivo esté apuntalado por las diferentes ciencias y disciplinas, y en el que puedan participar todos los segmentos de la sociedad.

37. Hacen falta nuevas formas de legislar sobre la economía digital que **garanticen la privacidad**

y la adecuada protección de los datos. Es más, deben fortalecerse valores deontológicos tales como los conceptos de «permitido», «obligatorio» y «prohibido» a la hora de navegar por internet e interactuar con robots.

38. **Las compañías deberían crear comisiones de ética y seguridad de las personas**, y sumarse a las ONG que bogan por establecer las mejores prácticas y los estándares para la implementación de la IA y la robótica en pro del bien común. Deben desarrollarse protocolos adecuados para garantizar la seguridad de las personas en el uso de la IA: tal es el caso de la verificación por duplicado de estas tecnologías por parte de equipos de diseño independientes. La realización de pruebas éticas y de seguridad de las personas —que, por ejemplo, evalúen el impacto social o detecten la presencia de prejuicios raciales encubiertos— debería ser un prerrequisito para el lanzamiento de cualquier software de IA. Debería considerarse la creación de comisiones ciudadanas externas que evalúen todas las tecnologías, incluso las militares, en forma recurrente y transparente

39. Dado que son quienes diseñan los dispositivos robóticos y de inteligencia artificial, los **científicos e ingenieros** son responsables de garantizar que sus inventos e innovaciones sean seguros y puedan emplearse para el bien común.

Declaración elaborada por Joachim von Braun (presidente de la PAS, con importantes aportes de los participantes de esta conferencia)

[1] Ambas Academias se encargaron de organizar una conferencia anterior (el 30 de noviembre y el 1 de diciembre de 2016), titulada «Poder y límites de la inteligencia artificial». La presente declaración se funda en el trabajo de esta conferencia de 2016 (incluyendo la declaración elaborada al término de la misma).