



Código	FPI-002
Objeto	Protocolo de presentación de proyectos de investigación SIGEVA UNLaM
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	04/09/2023

**Unidad Ejecutora:
Humanidades y Ciencias Sociales**

**Programa de acreditación:
CyTMA2**

**Título del proyecto de investigación:
“Filosofía de la Inteligencia Artificial”**

**PIDC:
Secretaría De Ciencia Y Tecnología**

**PII
Humanidades y Ciencias Sociales**

Directora del proyecto: Dra. Karina S. Pedace

Integrantes del equipo:

Docente- Investigadora: Mg. Ángela M. Menchón

Docente- Investigadora: Lic. Antonella Salvioli

Fecha de inicio: 1/1/2024

Fecha de finalización: 31/12/25

1-Cuadro resumen de horas semanales dedicadas al proyecto por parte de director e integrantes del equipo de investigación:¹

Rol del integrante	Nombre y Apellido	Cantidad de horas semanales dedicadas al proyecto	Graduado UNLaM	
			Si	No
Directora	KARINA SILVIA PEDACE	18		x
Docente-investigadora UNLaM	ANGELA M. MENCHON	8		x
Docente-investigadora UNLaM	ANTONELLA SALVIOLI	7	x	

2-Plan de investigación

2.1 Estado actual del conocimiento²:

Desde la antigüedad clásica, se ha sostenido la idea de que los seres humanos nos caracterizamos por ser animales con *logos*, es decir, con razón, con palabra. Nos concebimos a nosotros mismos como “animales racionales” y entendemos que lo que nos hace peculiares y nos distingue como especie biológica del resto de los animales y de las máquinas (i.e. autómatas, robots) es nuestra inteligencia, nuestra racionalidad y nuestra palabra. En efecto, durante mucho tiempo se ha pensado que solo los humanos podemos resolver ciertos desafíos (teoremas matemáticos, juegos como el ajedrez, o el Go), y se ha sostenido que hay conductas que solo los seres humanos podemos realizar (v.gr. escribir un poema, entablar un diálogo con otro ser humano). Tanto nuestra inteligencia teórica (i.e. nuestra capacidad para comprender el mundo y acumular conocimiento) como nuestras acciones (i.e. nuestra capacidad para resolver los desafíos prácticos que el mundo y la sociedad humana nos plantean) se han concebido como el resultado de nuestra capacidad de resolver problemas racionalmente.

A mediados del siglo XX nuestra concepción de lo humano fue puesta en jaque por el matemático y criptógrafo británico Alan Turing (1950) en el incisivo artículo “¿Pueden pensar las máquinas?”. En este trabajo, Turing, haciéndose eco de los desarrollos en lógica y en ciencias de la computación de la época, a los que él mismo contribuyó de manera notable, desafió la idea cartesiana que desterraba a las máquinas de la posibilidad de entablar un diálogo como si fueran un humano. El "juego de la imitación" –tal como Turing lo concibió– permitiría a una máquina (un autómata programado para entablar un diálogo con un humano) "engañar" al humano, confundiéndolo con otro humano. Así, la inteligencia artificial (IA) surgió bajo la premisa de construir un artefacto capaz de ejecutar conductas inteligentes indistinguibles de los patrones de conducta humana.

Este proyecto puede leerse canónicamente de dos maneras. Por un lado, de un modo menos ambicioso o estrecho (*narrow AI*) si se trata de buscar sistemas que realicen una tarea particular de forma inteligente. Alternativamente, puede ser visto de un modo más ambicioso o general (*general AI*), si se busca crear un artefacto que en todos los aspectos actúe (y piense) de una manera humana (como p.ej. Rachel, una de las replicantes de la película Blade Runner). Dado que este segundo proyecto parece seguir estando relegado a un horizonte remoto (más allá de las promesas que

¹ Incluir todos los integrantes del equipo de investigación, agregando tantas filas para cada rol de integrante del equipo de investigación como sea necesario.

² Sigo aquí algunos de los lineamientos centrales de Pedace, K., Balmaceda, T. Pérez, D. Lawler, D. Zeller Echenique, M. (2020). Caja de herramientas humanísticas, DL, ISBN 978-631-00-0120-3.

muchos encuentran en el auge reciente y disruptivo de las inteligencias artificiales generativas)³, en este plan de investigación nos centraremos en el proyecto más acotado, consistente en los sistemas que intentan replicar artificialmente ciertas capacidades humanas específicas.

Asimismo, en la literatura reciente hay una idea adicional muy provocativa. Hay buenas razones para pensar que la inteligencia humana no solo depende de nuestra dotación biológica sino también de los nichos culturales que el ser humano ha sabido construir, creando artefactos como el alfabeto y los smartphones para potenciar nuestras capacidades cognitivas (Clark y Chalmers 1998; Clark 1997; 2008). La creación del alfabeto, la notación arábiga de los números naturales y la creación de lenguajes artificiales en general para la lógica, la aritmética y la computación, son herramientas que, acopladas a nuestra dotación biológica, han permitido a los seres humanos realizar proezas inimaginables hace 40.000 años (o incluso hace 150).

Podemos entonces preguntarnos si no será posible generar con el desarrollo de la IA un nuevo nicho cultural que propicie que los seres humanos cambiemos de una manera drástica (aún no imaginada) nuestras capacidades cognitivas. ¿No estaremos ahora en una situación análoga a la que estábamos antes de la existencia del alfabeto? El desafío de combinar estos desarrollos de IA con las capacidades humanas existentes, potenciando las capacidades humanas (*human enhancement*), ha posibilitado ciertas líneas teóricas que hipotetizan que tal vez estemos ante la posibilidad de generar seres “transhumanos” (Lilley, 2013), es decir, seres que rompan las barreras de lo que la dotación biológica junto con el moldeado cultural actualmente existente nos permite hacer.

Como vemos, el alcance y la profundidad del impacto de las tecnologías en la vida cotidiana es difícil de ponderar, pero no deja de sorprendernos: se trata de desarrollos muy recientes que están permeando en tiempo récord a nuestra sociedad. El vértigo de las transformaciones vuelve difícil la reflexión detenida, pero ya comienzan a verse las consecuencias no deseadas de sus aplicaciones apresuradas: por ejemplo, los algoritmos que están a la base de la IA parecen estar reproduciendo y ahondando las inequidades del *statu quo* y ayudando a perpetuar las injusticias que vivimos a diario.

En efecto, los desarrollos que traen aparejados implican cambios significativos en las comunidades humanas, nuestro medio, nuestros comportamientos y nuestras maneras de concebirnos. En particular, **la inteligencia artificial conlleva importantes impactos en dimensiones relativas a la ética, la política, el medioambiente, el mundo del trabajo, la salud, la comunicación, la educación, la seguridad, entre otras.**

En este proyecto abordaremos **los sistemas de inteligencia artificial desde una reflexión filosófica** que parte de la asunción de que no es posible comprender cabalmente la tecnología sin la sociedad en la que se despliega, ni a la sociedad sin la injerencia de la tecnología. En consonancia con ello, se trata de un espacio que busca un equilibrio crítico frente a la dicotomía que tiene lugar entre la tecnofobia y la tecnofilia.

A tal efecto, este plan de investigación abrevará en literatura canónica y actualizada y en el análisis de casos desde diferentes perspectivas filosóficas.

2.2. Problemática a investigar:

Los desarrollos recientes en inteligencia artificial (IA) vienen teniendo un amplio impacto en la vida cotidiana. Muchas tareas se simplifican, surgen nuevas maneras de comunicarnos, desaparecen formas de empleo, convivimos con sistemas de recomendación que inciden sobre nosotros en diversos ámbitos fundamentales para la vida humana (i.e. cuestiones culturales, laborales, médicas, jurídicas, etc.).

Estos desarrollos impactan tanto en la imagen que tenemos de nosotros mismos, como en las prácticas cotidianas, y a la vez que generan beneficios no están desprovistos de riesgos. En efecto,

³ Entre las cuales el ChatGPT ha resultado uno de sus más conspicuos representantes.

surgen muchos **problemas éticos y políticos** tanto desde la instancia misma del diseño, así como en el uso de estos desarrollos en diversos ámbitos sociales tanto públicos como privados ¿Quiénes son los dueños de los datos personales que cargamos en las apps que usamos? ¿Es aceptable generar un sistema de puntaje basado en un algoritmo de aprendizaje automático que no sabemos cómo funciona realmente para decidir sobre situaciones de la vida humana como el otorgamiento de libertad condicional a un presidiario, otorgamiento de créditos bancarios, etc.? ¿Podemos confiar en una máquina así cuando nos da un diagnóstico médico? ¿Qué recaudos hay que tomar a la hora de diseñar estos sistemas para garantizar derechos humanos básicos? ¿Qué cuidados deben tomar las instituciones públicas al hacer uso de estos desarrollos? Allí donde estos desarrollos son técnicamente posibles ¿son deseables en clave del tipo de vida a la que aspiramos vivir? Y ¿cómo debemos concebir estas cuestiones desde un pensamiento situado en América Latina?

Por su parte, surgen **cuestiones ligadas a la dimensión epistemológica** tales como: ¿Cómo discernir un derivado meramente verosímil, como producto de un sistema de IA, de aquello que es verdadero? ¿Cómo abordar la difuminación de la distinción entre lo verdadero y lo falso? ¿Qué desafío puede representar para la integridad de la información y las democracias la propagación de *fake news* facilitada por la lógica algorítmica? ¿Qué capacidades críticas deben desarrollar los usuarios, los ciudadanos comunes, antes estos desarrollos?

Asimismo, se impone clarificar su impacto en la dimensión de la **antropología filosófica y medioambiental**. En efecto, resulta muy relevante elucidar de qué hablamos cuando hablamos de “inteligencia artificial”, en qué sentido este modo de “antropomorfizar” la tecnología resignifica la imagen que tenemos de nosotros mismos, el mundo que habitamos y nuestras relaciones intersubjetivas ¿Es plausible hablar de genuina “creatividad” por parte de sistemas de IA? ¿Qué tipos de extractivismos parece haber aquí en juego? ¿Qué consecuencias conllevan para nuestra autoconcepción y las condiciones materiales de nuestro mundo?

En este proyecto de investigación profundizaremos en estos tópicos y sus problemas concomitantes y trataremos de reflexionar desde una perspectiva filosófica acerca de estos nuevos desafíos técnicos, prácticos y teóricos que afrontamos en el siglo XXI.

2.3. Objetivos:⁴

El **objetivo general** del plan de investigación es **promover una perspectiva humanística de la IA** en un contexto en el que los diversos sistemas basados en dicha tecnología ponen en jaque la imagen que los seres humanos tenemos de nosotros mismos y que, consecuentemente, alteran las formas de vida de las sociedades humanas. Los desarrollos recientes conforman proyectos transdisciplinarios y polifónicos en los cuales la mirada filosófica se hace imprescindible para analizar y clarificar las dimensiones prácticas y teóricas involucradas.

Objetivos específicos:

1-Proponer un marco filosófico comprensivo para analizar los procesos de cambio que conlleva la implementación de la inteligencia artificial en el modo en que nos comprendemos como seres humanos, tanto a nivel personal como interpersonal.

2- Vertebrar mapas conceptuales sobre los procesos de diseño y uso de tecnologías que incorporan inteligencia artificial (IA) para analizar las dimensiones éticas, políticas, epistemológicas, estéticas y antropológicas en juego.

3- Examinar diferentes perspectivas en torno a la relación entre ciencia, tecnología y sociedad en el ámbito de la IA que permitan abordar problemáticas complejas desde una mirada filosófica.

⁴ Detallar objetivo general y objetivos específicos.

4- Indagar modelos de comunicación pública y difusión de la inteligencia artificial en la sociedad que permita a las/os ciudadanas/os comprender esta tecnología y sus diversos impactos en la vida cotidiana.

2.4. Marco teórico:

No todos los desarrollos de IA tienen los mismos algoritmos detrás, es decir, no todos ellos responden a un único modelo matemático. No vamos a detenernos en detalles, pero **al interior de nuestro marco teórico necesitamos mencionar algunas cuestiones generales para hacer explícitas algunas de las herramientas que proponemos utilizar para pensar humanísticamente la IA.**

Los primeros “programas” de IA estaban basados en la idea de “máquina de Turing”. Como sabemos, se trata de un algoritmo (una máquina ideal/lógica, no una máquina efectiva material) muy sencillo que incluye una serie de órdenes, codificadas en una tabla donde se especifican todas las funciones del sistema, o las reglas de transición. Así, una máquina de Turing es capaz de ejecutar una multiplicidad de tareas en la medida en que puedan aplicarse serialmente estas transiciones de un estado de la máquina a otro. En estos modelos podemos apreciar que las reglas que sigue el sistema reproducen cánones normativos acerca de qué se debe hacer ante cada información o situación dada, aplicando serial y recursivamente las reglas para llegar a una respuesta del problema planteado. Estas máquinas seriales muy potentes para realizar ciertas tareas tienen serias limitaciones para otras (ya desde sus inicios los famosos problemas del marco y de la explosión computacional señalaron restricciones a este modelo).

Pero el desarrollo de la IA se ha alejado rápidamente de estas máquinas seriales. La mayor parte de los desarrollos de IA que están ubicuamente presentes en nuestras vidas están basados en otro tipo de algoritmos que permiten extraer patrones a partir de grandes cantidades de datos (denominados usualmente “machine learning” o “aprendizaje automatizado”).

Estos mecanismos parten de una serie más o menos grande y más o menos estructurada de datos (“experiencias pasadas”), a partir de los cuales descubren patrones preexistentes en ellos, lo que les permite predecir qué va a ocurrir ante la aparición de un nuevo dato, o realizar tareas varias en función de la “experiencia” pasada acumulada, del “aprendizaje” realizado. El proceso de aprendizaje varía, puede ser guiado o automático (es decir, puede haber humanos que corrijan el avance en el aprendizaje –maestros– o puede aprender la máquina interactuando consigo misma), activo o pasivo (esto es, puede darse en medio de una constante interacción con el entorno que retroalimenta el proceso de aprendizaje, o puede ser puramente observacional, i.e. la máquina adquiere información antes de empezar a actuar), puede ser orientado a un fin (que la máquina aprenda a realizar una tarea específica) o simplemente que encuentre patrones inesperados entre los datos preexistentes (minería de datos –data mining–).

Los algoritmos pueden ser variados (Shalev-Shwartz y Ben-David, 2014) pero, en todos los casos, se trata de “educar” un programa para generar un modelo de un cierto dominio de la realidad, que le permitirá al sistema realizar las acciones apropiadas. Sin entrar en excesivos tecnicismos, siguiendo a Mittelstadt et al. (2016), **cabe advertir que hay al menos dos acepciones de “algoritmo”:**

- Desde su **caracterización formal** es considerado un constructo matemático a los efectos de lograr un propósito dado bajo ciertos suministros.
- Desde su **empleo concreto** alude a qué debe ser implementado y ejecutado para llevar a cabo una acción y obtener ciertos efectos.

Un algoritmo completamente configurado incorpora, entonces, **la estructura abstracta matemática que se ha implementado en un sistema para el análisis de tareas en un dominio particular.** En lo que sigue, cuando hablemos de “algoritmos”, **estaremos haciendo**

referencia a este segundo sentido, esto es, a un algoritmo realizado en un artefacto concreto, que forma parte de nuestro mundo material. La reflexión filosófica en torno a los algoritmos parece requerir crucialmente que atendamos a cómo son implementados y ejecutados en programas de computación, en software y en sistemas de información. En particular, la escena actual está dominada por una clase particular de algoritmos que toman decisiones determinando, por ejemplo, la mejor acción a llevar a cabo en una situación dada, la mejor interpretación de los datos, etc. Tales algoritmos aumentan o reemplazan los análisis y las tomas de decisiones por parte de seres humanos, a menudo debido al alcance o a la escala de los datos y de las reglas involucradas. No obstante, los algoritmos nos interpelan éticamente no solo en virtud de la escala de análisis y de la complejidad en la toma de decisión, sino que la opacidad del trabajo que efectúan y su impacto en nuestras vidas es también motivo de especial reflexión

Cabe señalar que los sistemas de aprendizaje automatizado se caracterizan por la capacidad de definir o de modificar reglas de toma de decisión de manera autónoma. Un algoritmo de *machine learning* aplicado a tareas de clasificación, por ejemplo, consiste típicamente en un aprendiz que produce un clasificador con la intención de desarrollar clases que se puedan generalizar más allá de los datos de entrenamiento (Domingos, 2012).

Como señalamos, el aprendizaje puede ser supervisado (vía inputs etiquetados a mano) o no supervisado (el algoritmo mismo define los modelos que mejor encajan para dotar de sentido al conjunto de inputs). En ambos casos, el algoritmo define las reglas de toma de decisión para manejar nuevos inputs. En consecuencia, el operador humano no necesita comprender la razón subyacente a las reglas de toma de decisión producidas por el algoritmo (Matthias, 2004).

Sin embargo, es un *desideratum* de los diseñadores de IA que la transparencia no desaparezca completamente en este proceso, dado que los algoritmos que son poco predecibles o poco explicables son difíciles de controlar, de monitorear o de corregir (Tutt, 2016). Como señalan Mittelstadt et al. (2016), la transparencia es tratada a menudo ingenuamente como una panacea para las cuestiones éticas y políticas que emergen a partir de las nuevas tecnologías.

De esta manera, **vemos que los datos de los que parte el sistema son clave para que el sistema aprenda**, y que lo aprendido está en una relación directa con los datos de partida que le proporcionamos y con las correcciones que se van haciendo a lo largo del proceso de aprendizaje. Así, si el sistema parte de un conjunto pasado de fallos judiciales injustos, generará nuevos fallos injustos; si parte de datos sesgados (por ejemplo, de datos acerca de cuáles de hecho son las preferencias de varones y mujeres en nuestra actual sociedad patriarcal) generará ofertas publicitarias sesgadas reforzando el sesgo original: ofrecerá escobillones a las mujeres y automóviles a los varones. También nos mostrará más frecuentemente las opiniones y las fotos de los amigos a los cuales les pusimos un “like”, que las de amigos a los que nunca les comentamos nada. Y reforzará nuestras antiguas opiniones mostrándonos las opiniones de quienes concuerdan con nosotros (sesgo de confirmación).

Los desarrollos que llevan incorporada IA están muy extendidos en nuestras sociedades actuales y tanto su diseño como su uso plantean una infinidad de cuestiones éticas, políticas, epistemológicas y antropológicas, muchas de ellas novedosas y urgentes. Es notable la velocidad con la que ciertos sistemas y aplicaciones que involucran IA se suceden y transforman nuestra vida cotidiana.

A través del **presente plan de investigación pretendemos ofrecer una serie de herramientas provenientes de la filosofía con vistas a generar un pensamiento crítico** en aquellas/os que de alguna u otra manera tengan que tomar decisiones que conciernen al diseño, desarrollo y/o al uso de IA: i.e. diseñadores, desarrolladores, empresas, gestores de

políticas públicas y usuarias/os en general, ciudadanas/nos que vivimos en este nuevo nicho cultural atravesado por la IA.

2.5. Hipótesis de trabajo

La **hipótesis principal** del plan de investigación es que **resulta necesario promover una perspectiva humanística de la IA** en un contexto en el que los diversos sistemas basados en dicha tecnología ponen en jaque la imagen que los seres humanos tenemos de nosotros mismos y que, consecuentemente, alteran las formas de vida de las sociedades humanas.

Hipótesis secundarias:

- 1- Resulta pertinente proponer un marco filosófico comprensivo para analizar los procesos de cambio que conlleva la implementación de la inteligencia artificial en el modo en que nos comprendemos como seres humanos, tanto a nivel personal como interpersonal.
- 2- Es de crucial importancia para nuestras sociedades vertebrar mapas conceptuales sobre los procesos de diseño y uso de tecnologías que incorporan inteligencia artificial (IA) para analizar las dimensiones éticas, políticas, epistemológicas, estéticas y antropológicas en juego.
- 3- Es de suma relevancia examinar diferentes perspectivas en torno a la relación entre ciencia, tecnología y sociedad en el ámbito de la IA a efectos de que permitan abordar problemáticas complejas desde una mirada filosófica.
- 4- La indagación de modelos de comunicación pública y difusión de la inteligencia artificial resulta de vital interés para que las/os ciudadanas/os puedan comprender esta tecnología y sus diversos impactos en la vida cotidiana.

2.6. Metodología:

El carácter de nuestra investigación es de naturaleza preeminentemente conceptual y consiste, principalmente, en un análisis crítico de los conceptos, tesis, argumentos y problemas mencionados en el plan de trabajo cuyo diseño se detalla más abajo. No obstante, el presente plan aspira, asimismo, a aplicar en debates contemporáneos en torno al impacto de la inteligencia artificial en la sociedad y esclarecer mediante herramientas propias de la disquisición filosófica sus implicaciones prácticas, ético-políticas, así como sus derivas *-inter alia-* epistemológicas y antropológicas.

En este sentido, se procederá al relevamiento y análisis de la bibliografía citada y del material nuevo que surja en el transcurso de la investigación. Se elucidarán los términos clave, las definiciones y las tesis principales

s. Se tematizarán los presupuestos de las diversas posiciones en las que abrevaremos y las consecuencias que se desprenden de ellas. A partir de las reflexiones resultantes, se evaluará la viabilidad de las hipótesis propuestas.

Asimismo, cabe señalar que, **a pesar de su impronta fundamentalmente conceptual, el presente proyecto aspira a que los resultados obtenidos tengan una aplicación que trasunte en el fortalecimiento del pensamiento crítico de las/os estudiantes y graduadas/os de nuestra Universidad y que dé sustento teórico a la reflexión y conversación pública** en torno a la incorporación de nuevas tecnologías en nuestra sociedad.

En lo relativo a la factibilidad del desarrollo del proyecto, cabe hacer constar lo siguiente. El Departamento de Humanidades y Ciencias Sociales, donde se radicará la presente investigación, cuenta con la infraestructura y los servicios necesarios para llevarla adelante. Por su parte, los recursos a los que podemos acceder vía la Biblioteca "Leopoldo Marechal" de la Universidad Nacional

de La Matanza (Sala de lectura silenciosa, Sala de lectura parlante –con 500 puestos y servicio wifi, la Sala Virtual dedicada a la consulta de recursos electrónicos, Sala de Referencia, Biblioteca sonora y la Sala Auditorio) conjuntamente con Núcleos de Acceso al Conocimiento (NAC) y el equipamiento disponible en el laboratorio de informática, dotan de factibilidad al desarrollo de nuestro proyecto que, tal como hemos adelantado, es de talante preeminentemente conceptual.

Finalmente, a los efectos de abordar la problemática de la que se ocupa el proyecto, nuestro **plan de trabajo está estructurado en las siguientes unidades analíticas:**

Unidad 1: Reflexiones preliminares desde la filosofía de la técnica. ¿Qué es la tecnología? Diferentes aproximaciones y definiciones. Ontología de artefactos. Agencialidad y valores desde el momento del diseño.

Unidad 2: Inteligencia artificial: discusiones contemporáneas acerca del ámbito de lo mental. ¿Pueden pensar las máquinas? El desafío de Turing: el juego de la imitación. “Inteligencia artificial” se dice de muchas maneras: inteligencia artificial general e inteligencia artificial estrecha. El desafío de las inteligencias artificiales generativas. Inteligencia artificial y creatividad.

Unidad 3: Inteligencia artificial y el “pecado original”: ni tan artificial, ni tan inteligente. Extractivismo planetario y extractivismo de datos. El costo de la conexión. El impacto en el mundo del trabajo: mano de obra, capital y tiempo. Automatización y gestión de los cuerpos. Seres humanos & robots.

Unidad 4: Inteligencia artificial y valores: neutralidad imposible. ¿Automatización de la desigualdad? Algoritmo en sentido estrecho y algoritmo en sentido amplio. Definición del problema. Los datos nunca son “datos”. Algoritmo y caja negra: opacidad y transparencia. El ser humano en el bucle. Privacidad y DDHH.

Unidad 5: Inteligencia artificial y sesgos. Distintos tipos de sesgo. El sesgo de tecnología y la delegación en las máquinas: ¿justicia digital? Los sesgos de género. Pensar la tecnología digital con perspectiva de género: cuando lo ético es político.

Unidad 6: Inteligencia artificial y los algoritmos en acción: videovigilancia y tecnología de reconocimiento facial. Seguridad, vigilancia y exposición. Armas de “destrucción matemática”: aumento de la desigualdad. *Fake news* y ¿amenaza a las democracias?

Unidad 7: Inteligencia artificial y lo humano en cuestión. Humanismo, Transhumanismo y Posthumanismo. Resignificación de la vida: “onlife” y resignificación de la finitud: posteridades digitales. El problema de la alineación de la IA. Una inteligencia artificial “centrada en lo humano”.

2.7. Bibliografía:

Unidad [1]

- Lawler, D. (2008) “Una incursión ontológica al mundo de los productos de la acción técnica” en *ArtefaCToS*, vol. 1, n.o 1, noviembre 2008, 4-17
- Mitcham, C., & Mackey, R. (Eds.). (2004). *Filosofía y tecnología: Edición española de Ignacio Quintanilla Navarro* (Vol. 201). Encuentro.

Bibliografía complementaria

- Parente, D.; Berti, A. y Celis, C. (Coords.) (2022). *Glosario de Filosofía de la Técnica*. La Cebra

Unidad [2]

- Descartes, R. (1980) *Meditaciones metafísicas*. Charcas, Buenos Aires. trad. E. de Olaso, T.

Zwanck (hay otras traducciones) Primera y Segunda Meditaciones

-Pedace, K., Balmaceda, T. Pérez, D. Lawler, D. Zeller Echenique, M. (2020). Caja de herramientas humanísticas. Recuperado de: <https://proyectoguia.lat/wp-content/uploads/2020/05/Caja-de-herramientas-Humanistas.pdf>

-Turing, A. (1953) "Los aparatos de computación y la inteligencia", en D. Hofstadter y D. Dennett. El ojo de la mente. Buenos Aires: Sudamericana.

Bibliografía complementaria

-Arte es Ética (2023). Decálogo de la falsificación y la sustitución de valor humano. Recuperado de: <https://arteesetica.org/decalogo-de-la-falsacion-y-sustitucion-del-valor-humano/>

Unidad [3]

-Couldry N. & U. Mejías (2019). El costo de la conexión: cómo los datos colonizan la vida humana y se la apropian para el capitalismo. EGodot. Ciudad Autónoma de Buenos Aires

-Crawford, K. (2022). Atlas de Inteligencia Artificial. Poder, política y costos planetarios, FCE. Ciudad Autónoma de Buenos Aires.

-Parikka, J. (2021) Una geología de los medios. Caja Negra

Bibliografía complementaria

-Pasquinelli, M & Joler, V. (2021). El Nooscopio de manifiesto. laFuga 25, pp. 1-20.

Unidad [4]

-Eubanks, V. (2021). La automatización de la desigualdad. Herramientas de tecnología avanzada para supervisar y castigar a los pobres. Capitán Swing

-Pedace, K., Balmaceda, T. Pérez, D. Lawler, D. Zeller Echenique, M. (2020). Caja de herramientas humanísticas. Recuperado de: <https://proyectoguia.lat/wp-content/uploads/2020/05/Caja-de-herramientas-Humanistas.pdf>

Bibliografía complementaria

- Renieris, E. (2023) Beyond data. Reclaiming Human Rights at the Dawn of the Metaverse. The MIT Press.

Unidad [5]

-Balmaceda, T., Pedace, K., Lawler D., Pérez D., Zeller M. "Pensar la tecnología digital con perspectiva de género" ISBN: 978-631-00-0119-7

-Hao, K. (2019a). Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo. MIT Technology Review, febrero 8, 2019.

- Pedace, K.; Schleider, T.; Balmaceda, T. (2023) "Inteligencia artificial y sesgos. El caso de la predicción algorítmica del embarazo adolescente en Salta". Revista Iberoamericana de Ciencia, Tecnología y Sociedad, vol. 18, Núm. 53.

Bibliografía complementaria

-Garapon, A y Jean Lassegue (2023) Justicia digital. Revolución gráfica y ruptura antropológica. Jusbaire.

Unidad [6]

-Balmaceda, T., Schleider, T. y Pedace, K. (2021). "Bajo observación: inteligencia artificial, reconocimiento facial y sesgos". ArtefaCToS. Revista de estudios sobre la ciencia y la tecnología, 10(2)

-Calvo, E. y Aruguete, N. (2020). Fake news, trolls y otros encantos. Cómo funcionan (para bien y para mal) las redes sociales. Siglo XXI

-O'Neil, C. (2018). Armas de destrucción matemática: cómo el big data aumenta la desigualdad

y amenaza la democracia. Capitán Swing

Bibliografía complementaria

- Scasserra, S. "Un paraíso de datos: cómo los paraísos fiscales del mundo se convirtieron en los centros de datos de la economía digital" Recuperado de: https://www.tni.org/files/2023-01/un_paraigos_de_datos_-_tni_-_web.pdf

Unidad [7]

Braidotti, R. (2020) El conocimiento posthumano. Gedisa

-Lawler, D. y otros "El transhumanismo a la luz de las teorías postcognitivas de la mente y el enfoque praxiológico de la tecnología" en Transhumanismo y tecnologías de mejoramiento humano Universidad Nacional Autónoma de México 2022

-Sadin, E. (2020) La Inteligencia Artificial o el desafío del siglo. Caja Negra

-Sisto, D. (2022) Posteridades digitales. Inmortalidad, memoria y luto en la era de Internet. Katz Ed.

Bibliografía complementaria

-Pedace, K. Balmaceda, T., Lawler D., Pérez D., Zeller M., "Natural born transhumans", Revista de Filosofía Aurora, 2020, 32 (55).

2.8. Programación de actividades (Gantt):

Actividades 1er Año	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
Relevamiento bibliográfico	X	X	x	x	x							
Seminario Anual	X	X	x	x	x	X	x	x	X	x	X	X
Elaboración, presentación y publicación de trabajos					x	X	x	x	X	x	X	X
Actividades con otros grupos de Investigación y reuniones con personalidades del exterior								x	X	x	X	X
Elaboración de informe académico de avance y rendición económica											X	X
Actividades 2do Año	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
Seminario anual, Elaboración, presentación y publicación de trabajos	X	X	X	X	x	X	x	x	x	x	X	X
Actividades con otros grupos de Investigación y reuniones con personalidades del exterior			x	X	x	X	x	X	x	x	X	X
Difusión de resultados finales, elaboración de informe académico final y rendición económica						x	x	x	x	x	X	X

2.9. Resultados en cuanto a la producción de conocimiento:

Conforme a los objetivos hechos explícitos, se espera aportar a la elucidación conceptual de aquello que hay en juego en la noción misma de “inteligencia artificial” y nuestros diversos modos de antropomorfizarla. Se espera aportar a la conversación pública en materia de aquello que implica el *desiderátum* de una “IA centrada en lo humano” tal como aquel objetivo al que se aspira v.gr. a través de las Recomendaciones sobre la ética de la inteligencia artificial de la UNESCO. Se espera contribuir a la concientización de cómo la inteligencia artificial está cambiando el universo de la toma de

decisiones colectivas y, en tal sentido, clarificar las dimensiones inherentes a su injerencia: ética, política, epistemológica, antropológica y medioambiental. Se espera, a partir de la reflexión crítica en torno a los diversos tipos de sesgos que inciden en el ciclo de vida de la IA, contribuir a la reducción de los prejuicios discriminatorios que gravitan negativamente en la población mediante el análisis y el control de la influencia de estos prejuicios en el diseño y desarrollo de algoritmos para los procesos de aprendizaje automático y la contribución al diseño de políticas públicas orientadas a contrarrestar esos efectos en ámbitos educativos, mediáticos y de formación cultural.

2.10. Resultados en cuanto a la formación de recursos humanos:

Tal como se ha hecho expreso, nuestro equipo está integrado por dos docentes-investigadoras de la UNLaM: la Mg. Angela Menchón y la Lic. Antonella Salvioli. En tal sentido, creemos que es de esperar que su participación activa en el proyecto propicie el desarrollo sus respectivas actitudes críticas y fortalezca sus saberes disciplinares.

Asimismo, creemos que es de esperar que la formación de recursos humanos redunde en una articulación estrecha con el resto de la comunidad académica, permita el establecimiento de vasos comunicantes más profundos con otros Departamentos de nuestra Universidad (v.gr. con el Departamento de Ingeniería e Investigaciones Tecnológicas, con algunas de cuyas referentes la Directora del presente proyecto viene forjando vínculos teóricos conjuntos) así como en el marco de la participación en actividades propulsadas conjuntamente con el *Centro de Estudios Avanzados* de nuestro Departamento de Humanidades y Ciencias Sociales.

2.11. Resultados en cuanto a la difusión de resultados:

Es de esperar que los resultados del presente proyecto impacten tanto a nivel de su difusión en el ámbito académico nacional e internacional, así como en materia de divulgación en clave de comunicación pública de la investigación, a través de diferentes formatos y lenguajes.

En efecto, tal como puede advertirse en los CVs SIGEVA UNLaM correspondientes a las integrantes: es de esperar que tengan lugar la publicación de artículos en revistas reconocidas en el ámbito en el que se inscribe el presente proyecto, capítulos de libros, ensayos, así como impartir conferencias, ciclos de lectura abiertos a la comunidad, intervenciones en medios de comunicación tradicionales como TV, radios y también mediante ciclo de podcasts (cf. vgr. “Sueñan los androides”) con la firme convicción de que resulta medular difundir los resultados obtenidos haciendo partícipes a quienes han de integrar una conversación pública polifónica en torno a la IA y su impacto en nuestra sociedad: academia, ONGs, Estados, decisoras políticas, etc.

2.12. Resultados en cuanto a transferencia hacia las actividades de docencia y extensión:

Los resultados obtenidos en esta investigación tendrán transferencia inmediata hacia las diversas actividades de docencia y extensión de quienes integramos el presente proyecto (cf. SIGEVA UNLaM CVs). Entre otras instancias, cabe consignar: la asignatura Filosofía, del C.E.FI., así como Metodología de la investigación en el ámbito de la Licenciatura de Relaciones Públicas de la Universidad Nacional de La Matanza, en la Facultad de Filosofía y Letras de la Universidad de Buenos Aires, en diversos cursos de grado y posgrado (tales como, por ejemplo, el Programa de Actualización en IA desde una perspectiva humanística, FFyL, Posgrado, UBA, cuya Coordinadora Académica es la Directora del presente proyecto), así como de otras universidades nacionales y del

exterior. A su vez, las actividades a desarrollar promoverán el intercambio con diversos centros de investigación que se consignan en los apartados siguientes.

El proyecto abarca una temática de gran actualidad e impacta en diversas áreas de la filosofía, (especialmente en la filosofía de la tecnología, filosofía de la mente, ética y antropología filosófica, *inter alia*) así como también en disciplinas científicas (ciencias de la computación, neurociencias, lingüística, etc.) y en la dimensión pedagógica a la hora de fortalecer experiencias innovadoras en el aula así como la alfabetización digital de la ciudadanía, como condición necesaria para una reflexión crítica en cuanto al diseño, desarrollo y uso de tecnologías digitales que, han de ceñirse a la adopción de una perspectiva de género y en el marco fundamental del respeto a los DDHH.

En efecto, la “inteligencia artificial” es central, en suma, para la comprensión de nuestra condición humana misma. Todos estos ámbitos de reflexión se verán beneficiados con los resultados de este proyecto, que se comunicarán -entre otro ámbitos- en diversos eventos académicos tanto nacionales como internacionales mediante ponencias y/o publicaciones en revistas y capítulos de libros especializados en las áreas en cuestión.

2.13. Resultados en cuanto a la transferencia de resultados a organismos externos a la UNLaM:

Asimismo, se transferirán resultados a organismos externos a la UNLaM en el marco que propician diversas esferas de intervención de las integrantes del presente proyecto tales como: v.gr. la Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación, CONICET, la REDBIOÉTICA de la UNESCO, la Red de Mujeres Filósofas de América Latina de la UNESCO, la Defensoría del Pueblo y la Casa ONU en Argentina, así como a diversas universidades nacionales y del exterior (Cf. SIGEVA UNLaM CVs).

2.14. Vinculación del proyecto con otros grupos de investigación del país y del exterior:

Es de esperar que nuestro equipo de investigación profundice los vínculos preexistentes con el Grupo de Investigación en Inteligencia Artificial, Filosofía y Tecnología (GIFT) radicado en el Instituto de Investigaciones Filosóficas, Sociedad Argentina de Análisis Filosófico, Unidad Ejecutora del CONICET; con el Grupo de investigación de FAMAF, Universidad Nacional de Córdoba y la Maestría en Tecnología, Política y Culturas bajo la dirección del Dr. Javier Blanco; el Observatorio de Impactos Sociales de IA, UNTREF, bajo la dirección de la investigadora Sofía Scasserra asociada al Transnational Institute; con el grupo de investigación integrado -inter alia- por las Dras. Flavia Costa, Noelia Billi, Paula Fleisner, Guadalupe Lucero y Claudia Kozak en el marco del Doctorado en Artes y Tecnoestéticas, UNTREF; el grupo de investigación que trabaja en el marco del Programa de Actualización en Tecnologías en el arte contemporáneo dirigido por la Dra. Jazmín Adler.

Asimismo, se espera estrechar lazos con quien ya hemos realizado durante 2023 un workshop sobre IA, agencia y valores y su grupo de investigación: Dr. Carlos Montemayor (San Francisco State University) y los grupos de investigación liderados por el Dr. Fernando Broncano (Universidad Carlos III, Madrid) y el Dr. Jesús Vega Encabo (Universidad Autónoma de Madrid), entre otros grupos del exterior.

2.15. Destinatarios:

Tipo de destinatario		Subtipo de destinatario ⁵	¿Cuál? Especificar	Demandante ⁶	Adoptante ⁷
Sector Gubernamental	Gobiernos	Del Poder Ejecutivo nacional			
		Del Poder Ejecutivo provincial			
		Del Poder Ejecutivo municipal			
	Otras Instituciones gubernamentales	Poder Legislativo en sus distintas jurisdicciones			
		Poder Judicial en sus distintas jurisdicciones			
Sector Salud		Hospitales, centros comunitarios de salud y otras entidades del sistema de atención			
Sector Educativo	Sistema universitario		x		UNLaM
	Sistema de educación básica y secundaria				
	Sistema de educación terciaria				
Sector Productivo	Empresas				
	Cooperativas de trabajo y producción				
	Asociaciones del Sector				
Sociedad Civil	ONG's y otras organizaciones sin fines de lucro		x		UNESCO-Red de Mujeres Filósofas y REDBIOÉTICA DE UNESCO
	Comunidades locales y particulares				

⁵ Marcar con una X

⁶ Demandante: entidad administrativa de gobierno nacional, provincial o municipal constituida como demandante externo de las tecnologías desarrolladas, que determina la necesidad del proyecto por su importancia social. Marcar con una X

⁷ Adoptante: beneficiario o usuario en capacidad de aplicar los resultados desarrollados (organismos gubernamentales de ciencia y tecnología nacionales o provinciales; universidades e institutos universitarios de gestión pública o privada; empresas públicas o privadas; entidades administrativas de gobierno nacionales, provinciales o municipales; entidades sin fines de lucro; hospitales públicos o privados; instituciones educativas no universitarias; y organismos multilaterales. Marcar con una X

3-Recursos Existentes⁸

Descripción/ concepto	Cantidad	Observaciones

4-Recursos financieros⁹

	Rubro	Año 1	Año 2	Total
Gastos de capital (equipamiento)	a) Equipamiento (1)			
	b) Licencias (2)			
	c) Bibliografía (3)	20.000		
	Total Gastos de Capital	\$ 0,00	\$ 0,00	\$ 20.000,00
Gastos corrientes (funcionamiento)	d) Bienes de consumo		10.000	
	e) Viajes y viáticos (4)	70.000	80.000	
	f) Difusión y/o protección de resultados (5)	30.000	30.000	
	g) Servicios de terceros (6)			
	h) Otros gastos (7)			
	Total Gastos Corrientes	\$ 100.000,00	\$ 120.000,00	\$ 220.000,00
Total Gastos (Capital + Corrientes)		\$120.000,00	\$120.000,00	\$240.000,00

Aclaraciones sobre rubros del presupuesto

1 Equipamiento: Equipamiento, repuestos o accesorios de equipos, etc.

2 Licencias: Adquisición de licencias de tecnología (software, o cualquier otro insumo que implique un contrato de licencia con el proveedor).

3 Bibliografía: En el caso de compra de bibliografía, ésta no debe estar accesible como suscripción en la Biblioteca Electrónica.

4 Viajes y viáticos: Viajes y viáticos en el país: Gastos de viajes, viáticos de campaña y pasantías en otros centros de investigación estrictamente listados en el proyecto. Gastos de viaje en el exterior: (no deberán superar el 20% del monto del proyecto).

5 Difusión y/o protección de resultados: Ej.: (Gastos para publicación de artículos, edición de libros inscripción a congresos y/o reuniones científicas).

6 Servicios de terceros: Servicios de terceros no personales (reparaciones, análisis, fotografía, etc.).

7 Otros gastos: Incluir, si es necesario, gastos a realizar que no fueron incluidos en los otros rubros.

4.1 Origen de los fondos solicitados

Institución	% Financiamiento
UNLaM	100%
Otros (indicar cuál)	-

⁸ Antes de confeccionar el presupuesto del proyecto, será necesario que el Director incluya en esta tabla si dispone de recursos adquiridos con fondos de proyectos anteriores (equipamiento, bibliografía, bienes de consumo, etc.) a ser utilizados en el proyecto a presentar, y además se recomienda consultar en la Unidad académica la disponibilidad de recursos existentes factibles de ser utilizados en el presente proyecto.

⁹ Justificar presupuesto detallado. Para compras de un importe de entre 6 y 60 módulos, (cuyo valor se regula de acuerdo con el Decreto 1030/16, Art 28) se requieren tres presupuestos. Superado ese límite, las contrataciones se realizarán según la normativa que rige las compras de la Universidad, todo de acuerdo con la Resolución HCS 143/2022.