



El riesgo está del lado de la crítica abstracta más que del retorno al empirismo ingenuo

The risk lies more on the side of abstract criticism than on a return to naive empiricism

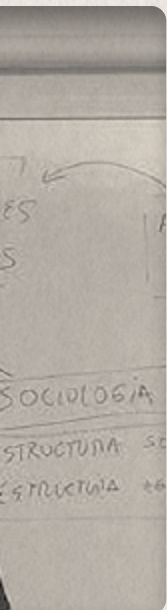
Entrevista a **Germán Federico Rosati** por **Beatriz Soria**, **Anabella Abarzúa Cutroni** y **Francisco Nicolás Favieri**

Recepción: 16/12/2024

Aceptación: 30/12/2024

.....
Germán Federico Rosati. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Altos Estudios Sociales (IDAES), Universidad Nacional de San Martín (UNSAM), Argentina.

✉ grosati@unsam.edu.ar



Beatriz Soria. Facultad de Ciencias Políticas y Sociales (FCPyS), Universidad Nacional de Cuyo (UNCUYO), Argentina.

✉ soriabeatriz@yahoo.com.ar

Anabella Abarzúa Cutroni. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Ciencias Humanas, Sociales y Ambientales (INCIHUSA), Facultad de Ciencias Políticas y Sociales (FCPyS), Universidad Nacional de Cuyo (UNCUYO), Argentina.

✉ a.abarzuacutroni1983@gmail.com

Francisco Nicolás Favieri. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Gabinete de Estudios e Investigación en Sociología (GEIS), Programa de Investigación y Estudios del Trabajo (PIET), Instituto de Investigaciones Socio-Económicas (IISE), Facultad de Ciencias Sociales (FACSO), Universidad Nacional de San Juan (UNSJ), Argentina.

✉ favieri@unsj-cuim.edu.ar

Palabras clave

Ciencias sociales computacionales, enfoques computacionales críticos. Modelos de lenguaje

Keywords

Computational social sciences, critical computational approaches. Large Language Models

RESUMEN

Germán Rosati, sociólogo e investigador del CONICET, comparte en esta entrevista sus perspectivas sobre el vínculo entre ciencias sociales y enfoques computacionales. Como director del *Diploma en Ciencias Sociales Computacionales y Humanidades Digitales* de la UNSAM, Rosati propone una reflexión crítica que integra teoría, empiria y técnica, destacando la necesidad de construir puentes entre el lenguaje sociológico y el computacional.

Abordamos temas como la reproducibilidad técnica en sociología, el rol del conocimiento técnico para fundamentar análisis críticos, y las tensiones entre las tradiciones teóricas y metodológicas frente a la transformación tecnológica. Rosati subraya que el valor de estas herramientas radica en cómo son orientadas por preguntas sociológicas relevantes, más allá de su mera aplicación técnica.

La entrevista es una invitación a re-imaginar la práctica sociológica en un contexto de Big Data e inteligencia artificial, equilibrando potencias y limitaciones para comprender la complejidad social actual.

ABSTRACT

Germán Rosati, sociologist and CONICET researcher, shares his perspectives on the relationship between social sciences and computational approaches in this interview. As the director of the *Diploma in Computational Social Sciences and Digital Humanities* at UNSAM, Rosati proposes a critical reflection that integrates theoretical, empirical, and technical dimensions, emphasizing the need to bridge the gap between sociological and computational languages.

The discussion addresses topics such as the importance of reproducible techniques in sociology, the role of technical knowledge in fostering critical analysis, and the tensions between theoretical traditions and methodological shifts in the face of technological transformation. Rosati highlights that the true value of computational tools lies in their application to sociologically relevant questions, beyond their technical implementation.

This interview invites a re-imagination of sociological practice in the context of Big Data and artificial intelligence, balancing the potential and limitations of these tools to better understand contemporary social complexity.

INTRODUCCIÓN

En esta entrevista, Germán Rosati, sociólogo, investigador del CONICET y docente de la Universidad Nacional de San Martín (UNSAM) nos comparte sus reflexiones sobre las discusiones, diálogos, tensiones entre las ciencias sociales y los enfoques computacionales. Desde hace unos años, Germán dirige el “Diploma en Ciencias Sociales Computacionales y Humanidades Digitales” de la UNSAM. En esta charla, además, nos proponemos repensar este vínculo sin perder de vista una perspectiva crítica que articule dimensiones teóricas, empíricas y técnicas. Además, es de interés poder examinar la necesidad de tender puentes entre los lenguajes de la teoría social y la computación, ya sea para revisar las formaciones de grado en las carreras de sociología pero también para pensar el futuro de la investigación social y profesional. Para ello es necesario reconocer las potencialidades que estas herramientas posibilitan pero también sus limitaciones.

En esta entrevista abordamos cuestiones esenciales como la necesidad de técnicas reproducibles en la sociología, y el papel del conocimiento técnico para desarrollar fundamentos críticos. Su perspectiva nos recuerda que la clave no está en la técnica *per se*, sino en cómo las preguntas sociológicas guían su aplicación.

La entrevista también explora temas relevantes como el impacto de los modelos de lenguaje en la investigación social, los desafíos de la docencia en tiempos de transformación tecnológica y las tensiones entre las tradiciones teóricas y metodológicas de la sociología. Este diálogo se convierte en una invitación para pensar los futuros (o los presentes) que las Ciencias Sociales pueden tomar en un mundo atravesado por el *Big Data* y la inteligencia artificial. Proponemos re-imaginar la práctica sociológica en tiempos de cambio pero también a aprender, criticar y construir nuevas herramientas para entender la complejidad social.

EXPERIENCIAS DE ESTUDIO Y TRABAJO: "HICE QUINIENTOS MILLONES DE ENCUESTAS"

Francisco: Para empezar nos podrías contar, ¿cómo es tu trayectoria académica y profesional? ¿Cómo fuiste acercándote al campo de las ciencias sociales computacionales?

Germán: Siempre es todo como racionalización *ex post*. Siempre todo tiene sentido después, pero bueno, hagamos como que funciona así. A ver, desde el tercer año de la carrera -más o menos- cuando cursaba el grado en sociología en la Universidad de Buenos Aires (UBA) empecé a trabajar como encuestador en la Encuesta Nacional de Gastos de los Hogares (ENGHo) y ese probablemente haya sido mi primer laburo estable. Estable no sería el término correcto pero sí con cierto grado de continuidad, y sobre todo el primero que tenía relación directa con lo que estaba estudiando. Entonces, entré a trabajar con lo cuantitativo, por decirlo de esa manera, desde muy temprano, ese fue el primer hito.

En ese momento, había una cátedra de Metodología de la Investigación Social que la dictaban personas vinculadas al Instituto Nacional de Estadísticas y Censos (INDEC), en ese momento Alejandro Giusti, Sebastián Bruno, Áurea Dias, Rodrigo Hobert y Ariel Lucarini y que funcionaba con un sistema de cohorte. Las tres metodologías giraban sobre un mismo tema y, en principio, la idea era que el grupo de estudiantes fuese el mismo durante los tres cuatrimestres. Esa materia fue un click, para mí. Tenía comentarios negativos: que era la peor materia de la carrera, por lo cual iba con las expectativas muy bajas. Y nada que ver, estaba muy bien, me voló la cabeza, digamos. Desde ese momento, me empezó a interesar el aspecto como metodólogo, antes de hacer ninguna investigación -con una soberbia total-, porque en el fondo nunca había hecho nada y ya me estaba preocupando por el cómo, bueno... cosas de estudiante.

Para la misma época, empecé a trabajar en el Programa de Investigación sobre el Movimiento de la Sociedad Argentina (PIMSA), en el equipo de Estructura Social, ahí comenzamos a laburar con fuentes cuantitativas, los censos. Creo que lo primero que hicimos fue un trabajo sobre de estructura social con el censo del 1991 y del 2001,. Primero trabajamos con tabulados publicados que pasamos a SPSS para procesarlos porque recién se estaban empezando a publicar los microdatos. Esa fue como la primera experiencia de investigación, también ya muy vinculada a lo cuantitativo.

Después, pasé a ser editor de encuestas en el Gobierno de la Ciudad, seguí trabajando como encuestador en varios operativos y después, hacia 2005 o 2006, empecé a trabajar en el Centro de Investigaciones en Estadística Aplicada (CINEA), que es como un centro de investigación que presta servicios de herramientas de datos de la Universidad Nacional de Tres de Febrero (UNTREF). Etré ya como analista y a partir de ahí todo lo que hice estuvo vinculado a encuestas y relevamientos cuantitativos más tradicionales (es decir, más habituales).

A fines de 2008 gané una beca de doctorado del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Mi tema de tesis fue el impacto de la mecanización de la cosecha de algodón en Chaco, Argentina, es decir, el paso de la cosecha manual a la cosecha mecánica. Esa fue la primera vez que hice entrevistas en profundidad, historias de vida, historias laborales en serio. Entrevistas diseñadas por mí, que eran para mi propia agenda y no para otros. Esa fue como la primera experiencia intensa con algo cualitativo, con recolección de datos cualitativos.

Luego de finalizar el doctorado empecé a trabajar y vincularme más con temas como aprendizaje automático y en Procesamiento de Lenguaje Natural (PLN)¹ y demás, en la Universidad Nacional de San Martín (UNSAM). En la UNSAM había una propuesta de proyecto... un tanto delirante, pero hermoso que consistía en armar una medidora estatal de rating en el año 2014. Era una propuesta que se vinculaba como parte de la Ley de Medios. En ese momento estaba la necesidad de una medición estatal que fuera una alternativa a IBOPE², que era (creo) la única medición de rating que existía en Argentina. Ahí es donde empecé a trabajar también con volúmenes de información mucho más grandes,

1 Se trata de un subcampo de la informática y la inteligencia artificial (IA) que emplea el aprendizaje automático para permitir que las computadoras comprendan y se comuniquen con el lenguaje humano. Se dedica al estudio y la creación de algoritmos y modelos digitales diseñados para procesar, interpretar y generar textos en lenguaje natural, es decir, aquellos escritos o hablados que utilizan el lenguaje humano tal como se emplea de manera cotidiana, incluyendo gramática, vocabulario y estructuras propias de una lengua.

2 Es una empresa de servicios dedicada a realizar mediciones de audiencia en los medios de comunicación en Argentina, de televisión, radio, publicidad exterior, entre otras.

sin llegar a ser Big data, y con datos muy diferentes, porque eso eran datos de eventos. La medición de rating funcionaba con un aparatito que graba el sonido de la televisión. En función de ese audio (y comparándolo con un audio máster en gabinete) se determina qué programa se está sintonizando en esa televisión. Eso viaja por línea celular y, obviamente, hay mil problemas y contingencias en esa transmisión y hay datos que “se pierden en el camino”. Parte del trabajo era limpiar esa información. Fue la primera vez que me encontré con ese volumen de datos y de errores. Errores, por su parte, en los que no era posible ir a un cuestionario a revisar los datos que faltaban. No era técnico en sonido, ni mucho menos, así que había cosas que no teníamos mucha idea cómo resolverlas y teníamos que laburar en conjunto con ingenieros e informáticos.

Trabajar con gente de ingeniería fue muy interesante en ese sentido. Ahí sí, probablemente empecé a trabajar con datos mucho más grandes que sesenta mil filas en una tabla, como puede ser una tabla de la EPH. De ahí surgieron otros proyectos vinculados al análisis de redes sociales y cosas por el estilo. Y ahí empecé a trabajar un poco con PLN con técnicas todavía muy rudimentarias.

Bueno, y después logré entrar a carrera en el CONICET, en el medio trabajé en el Ministerio de Trabajo. Pero bueno, más o menos eso. Pero siempre en la parte de análisis de datos, diría siempre en ese rubro.

Beatriz: ¿Hiciste alguna formación específica vinculada a las ciencias sociales computacionales? ¿Fue más autodidacta?

Germán: Al principio, fue autodidacta. Eran herramientas o técnicas que iba necesitando por cuestiones o requerimientos laborales. Entonces cuando empecé a trabajar ahí en el CINEA, en la UNTREF, había cosas que nos preguntamos ¿cómo procesar esto? ¿cómo en un índice? Eso suponía una parte autodidacta (buscar material, papers, aplicaciones) y otra que se daba charlando con el resto del equipo. Luego avancé en formalizar algunas cuestiones: hice la maestría en estadística ahí en UNTREF, empecé, pero no terminé por la maestría en *Data Mining* de la Austral.

Un lugar que no lo nombré recién y que probablemente sea donde más aprendí, fue cuando me tocó dar clases y coordinar los cursos de *Data Science* en *Digital House*... Fue como una especie de locura, medio que ahí, casi que aprendí a programar en *Python* en serio. Lue-

go, yo tiendo a hacer y aprovechar mucho los cursos *online*. Suelo hacer bastantes Moocs³. Siempre en función a algún tema puntual o para profundizar algún aspecto. Últimamente estuve haciendo algunos de LLMs. En mi opinión son un gran recurso.

EL “STARTER’S KIT”: LO MÁS IMPORTANTE DE LAS CIENCIAS SOCIALES COMPUTACIONALES SON LAS CIENCIAS SOCIALES

Francisco: ¿Cómo definís a las ciencias sociales computacionales? ¿Existe algún campo disciplinar propio? Si es así, ¿Cuáles serían esas contribuciones? Digo, las contribuciones de las ciencias sociales a la ciencia de datos. ¿Qué aportes haría? Y a todo esto, como para cerrar el combo. ¿Qué aportes haría también para la sociología?

Germán: Empiezo por el final, que para mí es lo más fácil. Me parece que los aportes de... llamémoslo, por ahora, a un “enfoque computacional” (con toda la ambigüedad que tiene el término) a la sociología (o a las ciencias sociales), me parecen evidentes en un montón de aspectos.

Primero, diría yo, la posibilidad de abordar temas, problemas o preguntas, que a lo mejor antes estaban planteadas en forma teórica (o en forma conceptual) y ahora tenemos chances de abordarlas de forma más sistemática y, sobre todo, de forma empírica. Y quizás llegar a una respuesta, no taxativa -aunque probablemente nada sea posible responder así tan taxativamente. En fin, creo que se abre esa posibilidad de atacar sistemáticamente muchos problemas de las ciencias sociales en general y de la sociología en particular. Me parece que ahí hay un punto, ¿no?. Después podemos discutir todo lo que quieran sobre *Big Data*⁴, los sesgos y demás.

3 Un MOOC (Massive Open Online Course) es un curso en línea masivo y abierto, diseñado para ofrecer formación a gran escala, accesible a cualquier persona con conexión a Internet, generalmente sin costo. Combina recursos como videos, lecturas y actividades interactivas, a menudo en plataformas educativas.

4 Se refiere al manejo, procesamiento y análisis de grandes volúmenes de datos que, por su tamaño, velocidad de generación y variedad, superan las capacidades de los métodos tradicionales de gestión de datos. Estos datos pueden provenir de múltiples fuentes, como redes sociales, sensores, transacciones digitales, dispositivos móviles, entre otros, y son utilizados para extraer patrones, tendencias y conocimientos valiosos para la toma de decisiones.

Efectivamente es una discusión importante. Pero no deja de ser cierto que hoy hay un montón de fuentes e información que uno podría utilizar para plantear -insisto-, problemas que son más bien los problemas clásicos, entonces sí, ahí hay un aporte, me parece.

Un segundo aporte que considero... lo voy a decir así muy brutalmente. Si estamos de acuerdo en que la Sociología o las Ciencias Sociales son efectivamente ciencias (y yo creo que lo son y no tienen que andar pidiéndole permiso a nadie) también sería un tanto necio negar que tenemos una serie de problemas en términos de la aplicación del método científico. Las causas son múltiples: son disciplinas jóvenes, el objeto es complejo, etc. y yo probablemente esté de acuerdo con todas las explicaciones, pero no deja de ser cierto que hay un problema en ese punto.

Uno de esos problemas, que está entre los más importantes, en mi opinión, es la falta de replicabilidad o mejor dicho reproducibilidad. En Ciencias Sociales no nos hemos preocupado -yo no creo que no se pueda-, pero soy más de la idea de que no nos hemos preocupado lo suficiente por hacer reproducibles los hallazgos de nuestras investigaciones.

Cualquiera que haya trabajado en investigación lo sabe, el problema es múltiple: va desde el investigador que se guarda los datos y no los quiere compartir -que es el caso más mezquino de la cuestión- hasta posturas epistemológicas que plantean que efectivamente la reproducibilidad no es posible. En ese *continuum* hay de todo, pero hay un problema ahí. Creo que el enfoque computacional, digamos, con todas las comillas del caso, plantea una diferencia potencial que va asociada a la producción de código de análisis, de procesamiento y modelado, a la liberación de fuentes de datos, al uso de datos abiertos. Me parece que puede avanzar o puede potenciar un poco esa falta de reproducibilidad que solemos tener en las Ciencias Sociales.

Me parece que esos son los dos aportes más importantes. Probablemente haya alguno que ahora me estoy olvidando, pero creo que esos son los aportes de los cuales nosotros nos podemos aprovechar, desde las Ciencias Sociales.

Siguiendo con la pregunta, el tema de si era un campo disciplinar o no. Yo pienso cada vez más que no. Para mí, lo importante de la Ciencia Social Computacional son las primeras dos palabras, reformulando, para mí no debería ser un campo disciplinar. Debería ser obvio

que en las Ciencias Sociales se usan estas herramientas, como se puede usar otras. No debería existir la necesidad, para mí, de un campo que sea Ciencia Social Computacional, como tampoco debería existir un campo que sea Ciencia Social Cuantitativa o Cualitativa.

En relación a las características que tiene, bueno... es ciencia social que usa técnicas computacionales y fuentes diversas, pero yo siento que para mí la ciencia social debería ser eso. Debería usar fuentes diversas, debería usar métodos diversos. Por eso digo, no sé si me cierra tanto pensarlo como un campo, porque creo que no debería existir.

Francisco: Con lo que comentaste recién ¿Cuál es tu perspectiva sobre el desarrollo de estas técnicas en Argentina?

Germán: Yo creo que se está empezando un poco a ver que hay posibilidades de abordar ciertas cuestiones, usar ciertas técnicas, abordar ciertos problemas. Me parece que hay algo que está empezando. De hecho, creo que hay por lo menos dos Diplomas que yo conozca que trabajan sobre estas herramientas desde la perspectiva de la ciencias sociales. Quiero decir, no una maestría en Ciencia de Datos que recibe gente de Ciencias Sociales. Eso también hay un montón, sino programas que se plantearon desde el lado de las Ciencias Sociales intentando abrir un espacio de formación en estas técnicas. Como todo, después hay que ver qué queda, ¿no? Porque también para mí, ahora hay como una explosión de cursos, diplomas, posgrados.

Francisco: Claro, ¿vos ves que hay como grupos o equipos de investigación, espacios que se van consolidando en esta línea un poco con eso?

Germán Rosati: Yo creo que... no sé si espacios, no sé si grupos de investigación, porque para mí el problema es que en el fondo no deja de ser un enfoque metodológico. Es decir, son herramientas. Entonces, no es que hay un equipo de investigación en Ciencias Sociales Computacionales, o al menos yo no lo conozco y no sé si tiene mucho sentido que exista. Hay algo ahí del orden de que lo podés aplicar a cualquier cosa, y en algún punto es la idea, es difícil hacer investigación centrado solamente en eso, en abstracción de cualquier problema concreto.

Lo que creo es que se están empezando a usar este tipo de herramientas, y que se están armando espacios,

sobre todo de discusión, de intercambio. Las últimas dos o tres Jornadas de Sociología de la UBA hubo una mesa al respecto. En las de San Juan también hubo una mesa, como que se está empezando a abrir como espacio de discusión. Y para mí eso está bien, pero otra vez, yo no sé si podés encuadrar un equipo de investigación en eso. En todo caso, de reflexión, de uso y de formación, diría yo, en este tipo de herramientas. Pero sí veo que comparado hace cinco años, me parece que hay una diferencia, incluso en la UBA también y me parece que muy de la mano del campo de NLP, ¿no? Me parece que ahí hubo algo que a todo el mundo le interesa, incluido yo. Con lo que mucha gente se engancha, porque permite, otra vez, plantear todas una serie de hipótesis sobre el lenguaje que antes siempre estaban planteadas en términos muy teóricos y permite ir bajando a tierra algunas ideas. Entonces, me parece que tiene como un atractivo en ese aspecto.

Francisco: En la línea de las Ciencias Sociales Computacionales ¿qué aspectos considerarás que son básicos para hacer Ciencia Social Computacional en la aplicación de estas estrategias?

Germán: Sí, por un lado, algo que uno da por hecho, es la formación en Ciencias Sociales. Es decir, tener en la cabeza una serie de problemas vinculados a la ciencia social, sino ¿cuál es la diferencia entre esto y lo que puede hacer una empresa que arma sistemas de recomendación, por decir cualquier cosa? No porque esté mal hacer eso, digo, pero ¿cuál sería la diferencia entre hacer investigación básica y producir ese tipo de herramientas? Entonces, para mí eso es central. O sea, el planteo de problemas de la ciencia social.

Después es una cosa que yo sí creo importante es una formación metodológica: saber operacionalizar conceptos, cómo poder pasar de una pregunta de investigación general y abstracta a algo que es operacionalizable, medible y abordable. Ese es uno de los problemas que tengo a ver, por lo menos, en parte de la formación de grado. Hay un salto entre la teoría, digamos, la teoría social y cómo se puede trabajarla en un proyecto de investigación concreto, ya sea investigación básica, aplicada, consultoría, en sector privado o público.

Y el tercer punto, sí es más técnico: contar con ciertas habilidades vinculadas a la cuestión de escritura de código, programación, estadística básica y algunas otras cuestiones relacionadas a eso. Para mí ese es como el... ¿cómo se dice? *starter's kit* para empezar.

Beatriz: Con esto que decís, ¿Cómo le ordenamos un poco el camino a un estudiante de sociología con estos temas? Empezar por dónde ¿cómo sería?

Germán: En términos de formación, yo diría que empiecen a pensar en formarse en programación y estadística, estadística en el sentido amplio. Incluye *Machine Learning*, en métodos cuantitativos en general/ Eso es lo que, creo, tiende a faltar en las carreras de grado en Ciencias Sociales y en general todavía no hay una oferta -me parece- en esa dirección. Esa pata es la que se van a tener que agenciar solos o vinculándose con equipos o con grupos que laburen en eso.

Tener una experiencia en la producción de datos, como encuestador no está mal, depende de cómo uno lo mire, ¿no? Está muy mal en términos de condiciones laborales, en términos salariales, pero ver un poco la “cocina” de la estadística en algunas cuestiones, me parece que está piola. Sobre todo en el Sistema Estadístico Nacional es muy esclarecedor, ver cómo funcionan ciertas categorías que, a lo mejor uno las tiene bastante claras en la teoría, y el contraste de ver qué en el campo no funcionan tan así, o el tipo de decisiones que se toman. En fin, desnaturalizar un poquito el mundo cuantitativo como si fuese totalmente aséptico y totalmente estandarizado. Me parece que transitar eso es interesante, insisto, más allá de la remuneración y de las condiciones de trabajo.

Para mí cualquier etapa de producción del dato, me parece que es interesante. Porque a veces, insisto, en las materias de metodología, uno da como una especie de esquema muy estilizado y muy cerradito del proceso de investigación: problema-pregunta-hipótesis-datos... Y está bien, porque hay que ordenar en términos pedagógicos y clarificar las etapas lógicas del proceso de investigación, cómo es el proceso de producción de la información. Pero en la realidad entre el momento en que se diseña el cuestionario, se hace el campo y se comienza el análisis, hubo una cantidad enorme de decisiones que se tomaron en el camino y de las cuales, a veces, uno no es consciente. En los relevamientos cuantitativos, que suponen un equipo de gente más o menos grande, me parece que es interesante tratar de entender, no idealizar el proceso de producción de información.

Entonces, pensar en formarse en programación, e insisto, en métodos cuantitativos en términos genéricos, es un buen primer paso. Luego, buscar algo que les interese para trabajar. Volviendo a lo que decíamos del

problema y de tenerlo en la cabeza dado que el meollo no es la técnica per se. Si uno no tiene algo en que aplicar lo que aprendió es complicado. Igualmente depende de cada uno, pero a mí me suele ser más estimulante pensar cuando tengo un problema en la cabeza sociológico, antropológico, lo que sea.

EL RIESGO A LA CRÍTICA ABSTRACTA Y LA RELACIÓN TEORÍA-MÉTODO

Beatriz: En relación a esto se ve que hay una revitalización de los estudios cuantitativos vinculados a los enfoques computacionales, pero también empiezan las alertas de un giro cuantitavista, como una cuestión más acrítica ahí. ¿Eso es algo que te preocupa? Pensando en quienes todavía creemos que se puede hacer Ciencia Social crítica, y bueno, si existe algo también ahí sumando a la Ciencia Social Computacional crítica, y digo, ¿te preocupa o es más decir, ‘bueno, como siempre, depende del uso que hagamos de los métodos y que no necesariamente está ligado a una perspectiva más epistemológica’?

Germán: Sí, yo soy más de la segunda, en principio. Enseguida matizo esto que digo. Pero empiezo por la segunda parte de la pregunta sobre la Ciencia Social crítica... A esta altura del partido, que me digan positivista, ya me resulta un elogio. Pero una técnica en el fondo es neutral. A veces se leen o se escuchan afirmaciones del estilo “*Machine Learning*⁵ es la CIA” y son un poco graciosas.

Las técnicas son técnicas. Hay un salto metodológico enorme entre una afirmación teórica, una técnica de investigación y el momento de la recolección de la información. Yo creo que hay muchas posturas que tienden a no ver esa mediación. Entonces para estas posturas lo cuantitativo es positivista, lo cuantitativo es la CIA. Y lo cualitativo es lo crítico. El problema es

⁵ Machine Learning (aprendizaje automático) es una rama de la Inteligencia Artificial que se centra en desarrollar algoritmos y modelos que permiten a las computadoras aprender y mejorar su rendimiento en tareas específicas a partir de datos, sin necesidad de ser programadas explícitamente para cada caso. Estos sistemas identifican patrones en los datos y los utilizan para hacer predicciones, tomar decisiones o realizar análisis, mejorando su precisión conforme reciben más información.

que algo parecido se da al revés. Porque ese es el otro punto, al revés pasa igual, lo cualitativo es metafísica y lo cuantitativo es ciencia verdadera. Entonces ahí me parece que hay una cuestión importante. Yo creo que lo primero es ponernos de acuerdo en que, en buena medida, las técnicas, son técnicas, y son de alguna manera “agnósticas” (no es un término que me guste del todo) respecto a las posturas teóricas que están de base, o las posturas teóricas que guían una investigación.

Aquello está tan arraigado, justo hoy me acordaba preparando este entrevista puntualmente, me acordaba de algo que había leído en una introducción a la edición alemana de “El oficio de sociólogo”⁶. Encima era una traducción al inglés, es decir que lo que uno lee es la traducción de la traducción. Ahí Bourdieu decía algo así como -yo me acuerdo porque lo tweeté-o: ‘Yo uso Análisis de Correspondencias Múltiples, porque es un método que está muy en línea con mis posturas teóricas sobre la teoría de campos y demás’. En mi opinión eso es un error, porque no hay teorías asociadas a métodos. En mi opinión, no hay problemas cuantitativos y problemas cuantitativos. Hay problemas de investigación de los cuales, a veces, tenemos el conocimiento suficiente como para estandarizar indicadores, entonces, los abordamos cuantitativamente. Hay problemas de los que no sabemos lo suficiente como para estandarizar la observación, y entonces, probablemente, ahí estemos más obligados a usar enfoques más cualitativos, más flexibles. Hay algo ahí que yo no termino de visualizar, ¿por qué se ve esa asociación tan fuerte entre teoría y método? Quiero decir: esa idea de que una teoría determina un método. Yo soy más de la segunda opción, en general, son técnicas y demás.

Ahora, dicho esto, eso no quiere decir que no se las puede usar mal, obviamente, pero el problema está en las hipótesis teóricas. Entonces, ¿puede haber ciencia social cuantitativa y de derecha, conservadora? Sí, obviamente, ¿cómo no? La hay, de hecho, ¿por qué no la va a ver? Ahora, el problema es invertir la causalidad y pensar ¿es de derecha porque es cuantitativa? o ¿es no crítica? Y por tanto, ¿es conservadora por ser cuantitativa y no por las hipótesis teóricas que la sustentan? Para hacer Ciencia Social Computacional crítica, en mi opinión, lo

6 En “El oficio de sociólogo” (1968), Pierre Bourdieu, Jean-Claude Chamboredon y Jean-Claude Passeron destacan la importancia del rigor metodológico y epistemológico en la sociología, abordando su práctica como una construcción científica sistemática y crítica.

importante es tener una teoría crítica. Y, en segundo lugar, todas las técnicas, obviamente, ¿no? Si no, es dar vuelta la causalidad, el orden lógico de las operaciones.

Y respecto al retorno a un empirismo medio ingenuo o demás, yo siempre le robo a Germani⁷ lo que dijo respecto al empirismo abstracto al estilo Lazarsfeld en el Prólogo a “La imaginación sociológica”⁸. Ahí plantea, algo así como: ‘Esto está muy bien, pero para Argentina no aplica, ¿no?’. Es un poco la operación de lectura de ese libro de Wright Mills y en este punto algo de razón tiene. Dice Germani, ‘los científicos latinoamericanos están más familiarizados con las críticas a las modernas técnicas de investigación que con las modernas técnicas de investigación en sí mismas’. Y yo creo que hay algo de eso. Insisto, no porque no se las pueda (y deba) criticar, no porque no se las pueda mejorar, pero primero hay que conocerlas. Si no, uno se queda una especie de crítica de la técnica muy abstracta al estilo de Eric Sadin⁹ o Byung Chul Han o Yuval Harari que es un acercamiento muy superficial al problema. En el caso de Sadin hay partes que son muy graciosas pero es porque no conoce lo que está criticando, y por eso hace una serie de afirmaciones que son técnicamente equivocadas, por decir lo menos... Hay muchísimos problemas, riesgos y demás que se le atribuyen a la Inteligencia Artificial, que en realidad son problemas que están desde mucho antes. Me parece que ya la crítica filosófica así, en esta instancia, llegó a un límite de lo que puede aportar, digamos. Desde Weber, Adorno, y compañía no sé cuánto más queda por decir al respecto.

Para mí el punto ahora es que hay que conocer las técnicas para poder eventualmente criticarlas, modificar-

7 Gino Germani (1911-1979) fue un sociólogo pionero de la sociología en América Latina. Su obra se centró en los procesos de modernización, la movilidad social y el autoritarismo en las sociedades latinoamericanas.

8 En “La imaginación sociológica” (1959), C. Wright Mills plantea la capacidad de conectar problemas personales con estructuras sociales más amplias, destacando la relación entre historia, biografía y sociedad para comprender el mundo social críticamente.

9 Éric Sadin (1973-) es un filósofo francés que critica el impacto de las tecnologías digitales en la sociedad, señalando los riesgos de la automatización y el control algorítmico. Sus obras analizan cómo estas transformaciones afectan la vida cotidiana y las relaciones humanas.

las o usarlas con una agenda crítica. Por eso, el riesgo que más preocupa en lo inmediato es el otro. Creo que es más riesgoso para las ciencias sociales el de una crítica abstracta que el del retorno al empirismo ingenuo, más en el ámbito argentino, latinoamericano... En todo caso, el riesgo del empirismo ingenuo viene más por el uso que otras disciplinas puedan hacer de problemas de las ciencias sociales que al revés, ahí hay algo de eso también, me parece.

Beatriz: Me quedé pensando en la primera pregunta, el primer bloque de preguntas sobre que donde hay las mayores transformaciones, es más que nada en el componente técnico. Pero, si no hay efectivamente algún impacto sobre el objeto propio de las Ciencias Sociales, si uno no redefine esto de que uno de los aportes tiene que ver con la posibilidad de contrastar empíricamente objetos que de otra manera quizás no hubiéramos podido hacerlo, o si quizás todavía siguen existiendo ciertos objetos privilegiados para la aplicación de métodos computacionales -que eso me parece más obvio- entonces ¿cómo hacemos? ¿cómo pensás esta cuestión?

Germán: Me imagino que es esperable que, cuando uno pueda abordar una pregunta, un problema, un objeto, o lo que sea, que antes no podía, es razonable que aparezcan preguntas nuevas, ¿no? Creo que, efectivamente, es así. Quizás podamos tener una aproximación a ciertos aspectos del discurso producido de forma más "espontánea", pero no más espontánea que una entrevista. Seguramente se van a abrir a preguntas nuevas y observables novedosos, también. ¿Cuánto...? Es una pregunta que vale la pena, porque ¿cuánto de eso es efecto de la técnica o de la fuente nueva? ¿y cuánto no? Me parece que es una pregunta que tiene sentido, porque la técnica en algún punto, es verdad, también redefine el objeto, pero una cosa es pensar que redefine el objeto o que abran preguntas nuevas o demás, y otra cosa es pensar que hay un campo disciplinar totalmente novedoso por la aparición de una técnica.

En particular, para mí, las preguntas de las Ciencias Sociales (las preguntas *realmente* importantes) son más o menos las mismas de hace ciento cincuenta años, con matices, con redefiniciones, pero siguen siendo más o menos las mismas. Y en algún punto, ¿por qué no lo serían? Para pocas hubo respuestas concluyentes, entonces, me parece que la aparición de estas técnicas, de estas herramientas y fuentes nuevas, creo que podrían conducir a refinar esas preguntas, hacerlas más específicas, eventualmente a plantear nuevas y a redefinir

objetos. Lo que quiero decir es que no me parece que el objeto sea sustancialmente distinto de las Ciencias Sociales, no es que hay un objeto nuevo.

Acá en algún punto estuvimos medio planteando, cuáles son los impactos y los efectos de estas técnicas sobre la Ciencia Social, pero después uno puede plantear la pregunta al revés. Diría que las Ciencias sociales en general pueden hacerlo, por ejemplo, ¿Cuáles son los impactos de la economía de plataformas? o es al revés, ¿Cómo la Inteligencia Artificial, cambia el objeto de la sociología? Entonces, ¿los mercados de trabajo de las aplicaciones son diferentes? ¿Sí o no? ¿por qué? Pero la pregunta es una pregunta clásica. Esa pregunta es muy pertinente y merece ser abordada pero, otra vez, no me parece que esa pregunta sea sustancialmente distinta a lo que se preguntaban los clásicos, hace cincuenta años, en condiciones nuevas y en otra etapa del desarrollo capitalista, claro.

Una cosa es que una técnica nueva permite abordar cosas que antes no se podían abordar empíricamente pero conceptualmente sí y que a partir de allí se abren matices, preguntas y demás. Otra cosa, es pensar que hay un objeto nuevo, es decir, que la digitalización genera un objeto completamente novedoso.

LOS "LARGE LANGUAGE MODEL" (LLM'S) Y LA AUTOMATIZACIÓN EN LOS PROCESOS DE INVESTIGACIÓN

Beatriz: También para pensar en lo que solemos enseñar, ¿no? El proceso de investigación tiene determinadas características, momentos o fases. En relación a este, pareciera que el principal cambio viene más por el lado instrumental o técnico, es decir, la alteración de ciertas fases o ciertos momentos del proceso de investigación. Parece que es bastante más claro en la parte de procesamiento y análisis, pero también se están empezando a ver transformaciones y mutaciones en la dimensión más puramente conceptual. Entonces, bueno, ¿cómo ha sido ese proceso para vos? Porque también es un aprendizaje en acto, me parece, muchas veces.

Germán: Sí, para mí, esto último es interesante. Creo que todavía son más bien tanteos, ¿no? Son intentos, son cuestiones que, me parece, se están empezando a explorar ahora, con lo cual tampoco sabemos hasta dónde. O sea, ¿cuánto de eso efectivamente va a tener impacto real, no? Sobre todo, y yo creo que ahí sí hay un salto cualitativo, hay una diferencia fuerte a partir

de los LLM¹⁰. Ahí hay un corte fuerte. Antes la “automatización”, operaba, creo, en dos aspectos. Uno, la aparición de fuentes nuevas y de datos masivos, y todo lo sabemos. Internet, la tecnología móvil, etc., “hacen disponible” datos -lo digo entre comillas porque no están totalmente abiertos-, pero bueno... digo, hacen disponible un montón de información que antes no estaba. Entonces, ese es un aspecto de la automatización.

Otro aspecto es la aparición de las técnicas de *Machine Learning*, del aprendizaje automático, que funcionan como formas de automatizar la búsqueda de modelos.. Con los LLM aparece una cosa nueva, creo, porque habrá que ver hasta dónde llega. Pero lo cierto es que parece abrirse la posibilidad de automatizar otras etapas del proceso de investigación. Por un lado, tareas como, por ejemplo la codificación de respuestas abiertas, o el subrayado de entrevistas, o todo lo que tiene que ver con trabajar con texto abierto, fueron siempre tareas que al tener que lidiar con texto abierto, tuvieron siempre un fuerte componente manual, se requerían equipos de codificadores, etc. Esas tareas parecen ser susceptibles de ser automatizadas en gran medida a partir de los LLMs. Quizás no pueda llegarse a una automatización total pero creo que nos encaminamos a un esquema de trabajo en el que la parte manual tiende a ser más revisar y verificar que sea correcto lo que hizo el LLM.

Podríamos pedirle a un LLM “codifícame tal cosa, codifícame ocupaciones”, para poner un ejemplo usando el CUIO¹¹. En general, ese tipo de tareas la resuelven con una buena performance y en un tiempo bastante menor que diez, veinte, treinta codificadores. Y en todo caso la parte manual es revisar que lo que hizo el “bicho” esté correcto, ¿no? Eso es una¹².

10 *Large Language Model (LLM)* son modelos de Inteligencia Artificial diseñados para procesar y generar lenguaje natural mediante redes neuronales entrenadas con grandes volúmenes de texto, permitiendo tareas como traducción, resumen y diálogo. Por ejemplo: GPT.

11 La Clasificación Internacional Uniforme de Ocupaciones (CUIO) es una herramienta para organizar las ocupaciones en una serie de grupos definidos en función de las tareas que comporta cada ocupación.

12 Un ejemplo de este tipo de tareas codificando orientaciones ideológicas de políticos en X puede verse en Törnberg, P. (2024). Large Language Models Outperform Expert Coders and Supervised Classifiers at Annotating Political Social Me-

Otra tarea que parece abrirse a la automatización son ciertas formas de captura de datos. ero no ya con *web scraping*¹³, y demás que, son más o menos, los bichos que escuchaban en redes sociales y o que *crawleaban* sitios web diversos. No, se empezaron a hacer pruebas de entrevistas en profundidad con *chatbots*. La persona interactúa con un *chatbot*, y el *chatbot* le repregunta. O sea, funciona como un entrevistador, no como un IVR¹⁴, no como un encuestador. Entonces ahí hay un par de *papers* que empezaron a hacer eso y están evaluando qué pasa con eso y cómo funciona¹⁵.

En el trabajo citado arriba, exploran un poco la experiencia de los encuestados mediante este *chatbot*, qué les pasa al interactuar con una máquina. Lo que se ve es que no parecen plantear una preferencia de un entrevistador humano por sobre la máquina, en términos generales. Habrá que pensar en la calidad de esa información recabada de esa manera. Pero ahí hay un pasito más. Es decir, es claro que aparece la posibilidad de automatizar parte de las etapas de recolección de datos, de procesamiento y de análisis. Pero además hay intentos de avanzar en la automatización en etapas previas: las etapas de planteo de problemas, de planteo de hipótesis, de diseño metodológico.

Hay algunos trabajos que lo que hacen es armar un entorno computacional al cual el usuario (investigador) le plantea una situación, por ejemplo, un juicio o una subasta con cinco escenarios, y el único *input* del modelo es la definición de la situación. Lo que el modelo

dia Messages. Social Science Computer Review, o(o). <https://doi.org/10.1177/08944393241286471>

13 Es una técnica para extraer automáticamente información de sitios web, transformando los datos disponibles en páginas HTML en formatos estructurados como tablas o bases de datos. Es útil para recopilar grandes volúmenes de datos de manera eficiente.

14 Interactive Voice Response (IVR) es un sistema automatizado de respuesta telefónica que permite a los usuarios interactuar con un menú mediante comandos de voz o teclas, para acceder a información o realizar acciones sin asistencia humana. Es común en servicios de atención al cliente.

15 Un ejemplo puede verse en Chopra, F. & Haaland, I. (2023). Conducting Qualitative Interviews with AI, CESifo Working Paper No. 10666, disponible en https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4583756.

hace es plantear hipótesis, relaciones causales, simular y analizar los resultados. En esos trabajos todos los ejemplos son ejemplos conocidos, porque en parte se trata de ver cómo funciona el entorno. Y pasan cosas como que el bicho simula una subasta de arte y entonces se observan cosas que ya son obvias, por ejemplo, cuando todos los compradores pujan por la obra, el precio sube y sucede lo contrario cuando baja. Si bien todos son resultados obvios y conocidos, el punto es que el bicho lo sacó de la nada. Solo con la definición de la situación generó todas las hipótesis, armó un esquema de simulación y analizó los resultados. Entonces, ahí uno dice 'yo no tengo tan claro cómo va a impactar eso'. Quizás termine siendo un fiasco pero bueno, hay una serie de cuestiones que antes no podíamos ni concebirlas como posibilidad y hoy sí.

¿Cuánto de eso serán “juguetitos” más o menos útiles para dominios muy acotados o cuántos efectivamente van a tener impacto real y significativo en la forma en que hacemos investigación? En las etapas de análisis y, precisamente, de información, me parece que hay elementos de sobra como para pensar que el impacto va a ser fuerte. Me parece que ahí no hay mucha vuelta que darle. Hay toda una serie de problemas potenciales que corresponden a la cuestión de los sesgos y demás, pero me parece que sí es clarísimo el impacto. En esta otra etapa no sé todavía. Pero de pronto se están empezando a probar, a testear ese tipo de herramientas.

Siempre dijimos, 'bueno, las preguntas las planteamos los seres humanos', las máquinas ejecutaban los pasos necesarios para dar una respuesta. Yo sería más prudente con ese tipo de afirmaciones absolutas (en ambos sentidos). Hace un par de años este tipo de cuestiones no eran ni concebibles. Ahora se están empezando a probar. Entonces, me parece que hay que estar atento a eso. ¿Y es un riesgo o no? No sé, justamente parte de toda esta situación un tanto vertiginosa es que no lo tenemos muy claro ¿eso nos va a liberar la cabeza para otro tipo de tareas? No sé. Yo espero que sí.

Beatriz: En las discusiones sobre la falta de transparencia en cómo los modelos de inteligencia artificial procesan los datos y generan resultados, la “cajanegrización”, ¿Te preocupa como sociólogo? ¿Qué priorizo? ¿Cómo balancear? Decir, por ejemplo, 'pierdo control y transparencia sobre el proceso de investigación pero ganó escalabilidad' por ejemplo ¿esa tensión está presente?

Germán: Sí, obvio, que sí. Sobre todo porque uno es víctima inconsciente del sesgo de confirmación, ¿no?

Si uno ve que las cosas funcionan bien, en el sentido de lo que uno está queriendo, está tentado automáticamente a asumir que esto funciona perfectamente. Yo creo que hay una cuestión a pensar ahí y que ya existe desde hace un tiempo, ¿no? Antes de los LLMs también la cuestión de la “cajanegrización” existía. Me parece que hay una cuestión a tener en cuenta.

Con la aparición de estos modelos enormes se plantean una serie de problemas que hay que tener en cuenta. Por ejemplo, no es un dato menor que hay unas cuatro o cinco empresas en el mundo con los recursos físicos, humanos, de hardware, etc que pueden entrenar desde cero un LLM. Esto tiene muchas aristas pero hay una que se vincula con algo que planteamos más arriba: la reproducibilidad. Si mañana a *OpenAI*¹⁶ se le ocurre pasar del GPT-4 al GPT-5.1, lo cierto es que no tenemos ninguna garantía de que lo que corrimos en GPT-4 va a tener los mismos resultados en este modelo nuevo, incluso en algo tan acotado como codificar una variable.

Estos tipos de modelos son opacos: no podemos saber bien qué es lo que está pasando adentro. A su vez, muchos de esos modelos son propietarios. Es decir, hay que pagar para usarlos. Y muchos, además, son cerrados en el sentido que no se puede acceder a sus parámetros. No tenemos forma de acceder a GPT, no se puede acceder a los parámetros del modelo. Todo esto atenta contra la reproducibilidad.

Eso cambió un poco, a mediados del año pasado, cuando Meta, que era la que se estaba quedando atrás en toda la “carrera armamentística” de estas herramientas, saca su propio LLM y lo “libera”. Lo hace libre para uso comercial, para uso académico y, además, lo libera en su totalidad: es decir, que hoy podemos bajar Llama (el modelo de Meta) y correrlo localmente (si disponemos de la infraestructura adecuada). No necesariamente tenés que usar un servidor de ellos. Entonces eso fue medio una pateada de tablero. Por que parte del problema de reproducibilidad queda solucionado. Ahí cambió un poquito la cosa, pero bueno, el riesgo está latente. Entonces, insisto, me parece que se hace todavía más necesario el saber o los saberes que la ciencia social tiene justamente para evaluar esas cuestiones. Los modelos en buena medida son cajas negras, los modelos, sobre todo, los LLMs y lo que nos tenemos que dar son estrategias para evaluarlos, pero no en términos de...

16 Es la empresa, fundada en el año 2015, creadora de ChatGPT.

¿Son “el mal”? ¿Son Skynet?. Me parece que es más productivo pensar en evaluaciones concretas asociadas a tareas concretas. Si yo estoy usando este modelo para esta tarea, bueno, **necesito tener criterios para evaluar** cómo funciona en esa tarea. A eso me refiero. Y antes de eso, ¿es este modelo el más adecuado para esta tarea que yo estoy queriendo realizar o no? A eso me refiero con evaluar. No a la crítica de la técnica, que se puede hacer, pero me parece que corre por otro canal.

LA CUESTIÓN DE LOS LLM'S Y LA ENSEÑANZA EN LAS CARRERAS DE GRADO

Anabella: Volviendo un poco a lo que planteabas de que los seres humanos somos los que hacemos las preguntas y pensando con melancolía en eso, en qué enseñar en el grado, con la irrupción del chat GPT y cómo incorporarlo a las clases sin ser conservadores ni prohibirlo, porque de todos modos lo van a usar. Tomando lo que vos decís de los criterios para evaluar los modelos ¿cómo integrarías vos el uso de estas herramientas?. Un poco la línea nuestra es decir, ‘bueno, hay que ser expertos en un tema para poder evaluar las respuestas que da el modelo y a la vez hay que saber preguntar y adquirir esa capacidad de preguntar’. Eso es lo que más o menos intentamos hacer en el grado, pero bueno ¿qué pensas vos de eso y qué te parece que sería útil enseñar?

Germán: Pues ahí yo separaría los usos posibles de estos “bichos”. Por un lado, está el que uno se imagina, que todo el mundo más o menos en este ámbito hace o está por hacer: usarlo como asistente para traducciones, para re-escrituras de texto, etc. Ese tipo de usos creo que es, en principio, el menos riesgoso. Y me arriesgaría a decir que es un uso muy productivo para el nivel de grado. Luego, está la cuestión de que es un uso medio inevitable, entonces, uno puede hacer como que estas cosas no existen. Está bien, puede fingir demencia todo lo que quiera, pero la verdad es que existen y se usan.

Otro uso, me parece tiene más que ver con el miedo de buena parte de los docentes... es decir, ¿cómo hago para garantizar que no le pidan a GPT que le escriba el examen o la monografía? Bueno, no sé. Me parece que esa es una pregunta mal planteada, porque es al revés, somos nosotros los que tenemos que diseñar un formato de evaluación que sea robusta a esas herramientas. Y en todo caso, pensar cómo podemos utilizar todas las potencias, incluso los problemas que tenés sobre las herramientas para tomarlo como parte del diagnóstico de evaluación.

Yo recuerdo a docentes (muchos de los cuales fueron mis docentes) que enseñan materias de teoría sociológica. Y muchos exámenes eran ‘¿Qué es la plusvalía?’ por poner un ejemplo. O ‘¿Qué es un hecho social?’. Y lo cierto es que esas preguntas ya eran “vulnerables” a la tecnología desde que existe Wikipedia (o Encarta para los más veteranos). Es cierto: en estos casos era más fácil detectar si alguien estaba copiando y pegando un texto. Ahora es un poco más difícil. Pero lo que quiero decir es que quizás el problema está en la pregunta que hacemos, no en la herramienta. La cuestión ahí son las consecuencias pedagógicas, que a mí un poco se me escapan, porque, la verdad, no es algo que yo maneje.

En mi caso, en el tipo de materias que doy, creo que es lo mejor que me pudo pasar. De hecho, yo suelo alentar el uso de GPT (o algún otro LLM) para escribir o para corregir código de R o Python. Por otro lado, si yo lo hago ¿por qué les voy a decir que no lo hagan? Bueno, si fueran carreras de programación, quizás no sea la mejor herramienta, porque es probable que el código que produce no sea el mejor código. Es probable que pueda escribirse de una manera más elegante, más performante. Y yo no lo discuto. Pero bueno, eso por un lado.

Después están todos los usos ya “fuera del aula” y en términos de investigación. Si la decisión es usar algún LLM entonces me parece que nociones mínimas de lo que se llama ingeniería de *prompts*¹⁷. En algún momento, hay que pensar cómo hacer para introducirlo en la formación de grado. No sé si como parte de metodología. La verdad es que no tengo claro cómo sería el diseño curricular, pero en algún lado va a tener que estar y lo mismo, creo, en relación a rudimentos en PLN. Quizás no para entrenar modelos, porque yo creo que todo lo que antes hacíamos con *Topic Modeling*¹⁸, se lo van a terminar comiendo estos “bichos”, y lo van a hacer mucho mejor. Yo creo que vamos hacia eso, pero sí para tener alguna idea de lo que está pasando por atrás.

17 La ingeniería de prompts (prompt traducido al español refiere a indicación, instrucción o solicitud) consiste en diseñar entradas claras y optimizadas para guiar modelos de lenguaje a generar respuestas precisas y relevantes. Se basa en iterar, ajustar y proporcionar contexto para mejorar los resultados.

18 El Modelado de Tópico (Topic Modeling) es una técnica de NLP que identifica automáticamente temas ocultos en grandes conjuntos de texto no estructurado. Es útil para descubrir patrones y categorizar documentos de manera no supervisada.

Por más que uno no conozca el detalle y el mínimo de estos bichos, pero sí es necesario entender ciertos aspectos de su funcionamiento. Incluso para evitar algunas discusiones medio bizantinas como la pregunta sobre si son inteligentes o razonan. En lo personal, yo creo que algún nivel de razonamiento hay, porque hay pasos y encadenamiento de pasos para una respuesta. No creo que haya inteligencia. Pero en todo caso, esta discusión importa poco en el uso cotidiano. Pero volviendo a lo anterior... yo creo que la ingeniería de *prompts* junto con algunos rudimentos en programación y NLP van a ser necesarios porque en el transcurso de una investigación real es altamente probable que la interfaz de chat sea suficiente.

En algún momento preguntarle al chat GPT, tal y como está, no va a alcanzar. Porque eso tiene más que ver con el uso como asistente del que hablábamos antes. El uso de las APIs¹⁹ va a haber que introducirlo. Eso supone dar alguna noción de programación, dar alguna noción de lo que es una API. Me parece que, en algún momento, va a haber que salir de la interfaz de chat como única forma de interactuar con estos modelos y pasar a un entorno más programático. Eso supone un poco de programación, e ingeniería de *prompts*, que me parece que va a haber que incorporarlo en algún momento. Insisto, no sé en dónde, si en metodología. No sé si vamos a dejar de dar estadística y vamos a pasar a dar ingeniería de *prompt* -no creo que eso sea deseable- pero sí se me ocurre que esos son los requisitos mínimos, digamos, como para poder hacer un uso mínimamente responsable y útil de estas herramientas.

Beatriz: Por otro lado, consultarte, estamos en una crisis de las carreras de sociología fuertísima en términos de matrícula, si quizás no hay una oportunidad de aire con la incorporación de estas formaciones.

Anabella: En Ciencia Política, también.

Beatriz: Te hemos escuchado hablar de la noción de “traductores”, es decir, tener la posibilidad de, por lo menos, dialogar en estos dos universos distintos y que nos entendamos que la comunicación sea posible,

19 *Application Programming Interface* (API) es un conjunto de reglas, protocolos y herramientas que permite a diferentes aplicaciones o sistemas comunicarse e intercambiar información. Actúa como un intermediario que facilita la interacción entre software, definiendo cómo solicitar y transmitir datos.

ese sería el punto. Pero contanos sobre este concepto, ¿cómo lo entendés? Después pensar, bueno, esto como una reflexión tuya, ¿cuán refractario o cuán complejo es ir a esa formación de analistas de datos sociales, cuando los y las propios docentes estamos formándonos en eso? Entonces, ¿apostamos a una formación más democrática -estamos todos y todas en un punto aprendiendo- o seguimos con las formaciones académicas tradicionales, donde el lugar del saber queda un poco más diluido? Estoy pensando sobre todo en la formación de grado.

Germán: Está bien el punto. A mí me parece que son dos cuestiones que son relevantes. La idea de “traductor” es más una impresión (y tampoco es algo que se me haya ocurrido a mí), pero sí me parece algo que es importante, sobre todo en relación a la idea de interdisciplina y los problemas asociados.

Cualquiera que le haya tocado trabajar en algún equipo mínimamente diverso en términos de disciplinas, se da cuenta que no alcanza con poner computólogos, físicos, sociólogos, antropólogos, para que funcione. La interdisciplina no ocurre por mera aglomeración disciplinaria. No hay emergencia ahí, no hay auto-emergencia. Esta idea “traductor” apunta a pensar en un perfil de personas que hablen los dos idiomas, que hablen, que conozcan la teoría social y los problemas de las ciencias sociales, que estén formados en los métodos de las ciencias sociales, pero también que estén familiarizados con el “mundillo” *machine learning*, inteligencia artificial, estadística, computación, pongámosle el rótulo que uno quiera. Eso debería, creo, facilitar justamente el trabajo común. Y ahí sí, el trabajo, inter, trans -no sé cómo llamarlo-, disciplinario, que haya diálogo entre las disciplinas en serio, sino es como que no va a funcionar. Para mí eso vale para los dos lados.

Yo también creo que los computólogos tendrían que aprender teoría social y metodología, pero mientras no pasa, bueno, nosotros podemos operar sobre este polo de la relación. Me parece que eso es importante, pensar en la posibilidad de poder dialogar y trabajar en equipos interdisciplinarios, y eso va más allá de la academia. En la academia y la investigación básica, seguro pero en un entorno laboral no académico -profesional- también.

A mí me ha tocado trabajar en consultorías y demás, con otras disciplinas, y en general, aunque no había traductores, por lo menos había un intento de buena

predisposición y un intento sincero de dialogar entre disciplinas. A veces, eso no ocurre y lleva a que se rompan equipos, o que proyectos que tendrían que ocurrir no ocurran. Me parece que la formación en estas técnicas es ineludible. No es solo para llevarse “bien” con otras disciplinas, sino me parece que enriquecería la propia disciplina también. No es una mera cuestión reactiva, creo.

Respecto a la segunda, para mí no es un problema, de hecho, hay muchos estudios que muestran que cuando hay mucha distancia -distancia pedagógica- en términos de conocimiento, práctica, saber, etc., entre un docente y estudiantes, los resultados son peores. Porque cuando alguien ha trabajado mucho tiempo en un tema o una técnica o lo que sea, hay ciertas relaciones conceptuales, lógicas, etc que ya están “automatizadas”. Esa persona no está repensando todo el tiempo en..., no sé, por poner un ejemplo, en cómo tenés que hacer un *for loop*²⁰. Ahora bien, un estudiante que está llegando, está obligado a pensar eso porque, probablemente, es la primera que se lo tope. Entonces, esa distancia en el fondo, termina siendo contraproducente porque ese docente experimentado ya realiza ese proceso casi inconscientemente, sin pensar demasiado en sus implicancias. No tiene que re-veer el proceso cada vez que ocurre.

En cambio, cuando la distancia -es horrible el término-, cuando la distancia pedagógica entre docente y estudiante es menor, en general tiende a funcionar mucho mejor el diálogo pedagógico y demás. Hay varios laburos de eso, muy vinculados a la enseñanza de tecnología, de disciplinas tecnológicas. Entonces, yo no veo un problema grave en esa especie de distancia corta, no sé cómo llamarlo, ¿no? El modelo de clase magistral hace años que está perimido, para mí. Escuchar a alguien tres horas hablar no sé cuán productivo puede ser, digo tres horas, no cuarenta minutos.

Lo otro vinculado a la cuestión de los “docentes refractarios”, es como una nueva categoría, un nuevo tipo social, ¿no? Existen, es así, yo creo que efectivamente pasa, no veo cómo resolverlo. Si hay un “docente re-

fractario” que da teoría social ¿qué les hará hacer? ¿un examen presencial? ¿les confiscarías los celulares durante el examen y los corregirá a mano? Bueno, qué sé yo, está bien. Nosotros nos formamos un poco en esa escuela, y tampoco fue tan mal eso.

Ahora, sí me parece que hay un problema en instancias más vinculadas a la práctica (académica o profesional), por ejemplo, a la producción de tesinas. Ahí sí, me parece que puede haber un problema. También en instancias de taller de tesis, puede haber una cuestión, porque efectivamente, si un estudiante se topa con un director de tesis o un docente de taller que es refractario... creo que allí puede haber un problema porque se está cerrando todo un arsenal metodológico antes de empezar... Yo desconozco cómo resolverlo, si no es con algún tipo de pauta institucional más o menos clara en esa dirección.

Beatriz: En relación a lo anterior y en parte con el debate sobre la matrícula ¿En la Diplomatura ha crecido?²¹

Germán: Bueno, este año sí, contrariamente a todo lo que nos imaginábamos. De hecho, aumentó incluso un poquito respecto al año anterior. Habrá unos quince o veinte estudiantes más. Después siempre hay un desgranamiento, obviamente, en la cursada, pero sí calculamos que van a terminar unas ochenta o noventa personas, que es más o menos la misma cantidad que el año pasado.

Bueno, ya no quiero abrir otra arista, pero hay cierto formato de formación que está un poco en crisis. Mucha de la gente que se acerca al Diploma tiene intenciones muy serias de formarse metodológicamente en este tipo de herramientas pero no quiere encarar dos años o tres años de cursada y más la escritura de una tesis de maestría, doctorado, lo que sea. Y en algún punto algo de sentido tiene esa intención, digamos.

Anabella: En el grado también, los estudiantes ya no quieren estudiar, estudiar con un cierto orden, una materia atrás de otra, una unidad detrás de otra, contenido que muchas veces les parece arbitrario.

Germán: Para mí el caso del grado es diferente, porque en esa instancia sí es muy necesaria cierta cons-

20 Un *for loop* es una estructura de control en programación que se utiliza para iterar sobre una secuencia (como listas, rangos o cadenas) y ejecutar un bloque de código repetidamente por cada elemento. Es ideal para recorrer colecciones o repetir tareas un número definido de veces.

21 Nos referimos al “Diploma en Ciencias Sociales Computaciones y Humanidades Digitales” que dirige Germán Rosati en la UNSAM desde 2021.

tancia. En el caso del Diploma hablamos, en general, de gente que ya está recibida, ya está formada, incluso transitó alguna formación de posgrado, en algunos casos, o la está transitando. Y lo que quiere es adquirir un conjunto de habilidades vinculadas al análisis de datos con una perspectiva “moderna”, por decirlo de alguna manera. No es un posgrado y es cierto que se abre una disyuntiva allí. Mi postura es mantenernos en ese formato “no posgrado formal” y que eventualmente después se pueda articular con otros posgrados más formalizados. De hecho, se han reconocido créditos en el Doctorado de la UNSAM y en algunas maestrías por cursar el Diploma. Pero creo que este formato nos sirve. En el nivel del grado creo que es necesaria una formación mucho más de mediano o largo aliento.

Anabella: Tiene que haber una sistematización, pero digo que cada vez les parece más arbitrario, ¿no? Un programa cerrado de contenidos.

Germán: Es verdad que suele haber en muchas carreras de grado poca articulación entre las materias... es como que pusimos los títulos y después el contenido. Armamos los vasos y cada quien le puso el líquido que quería, ¿no? Pero creo que en todo caso supone la necesidad de un mejor diseño curricular y no un achicamiento de contenidos.

UN POCO DE FUTURO PARA EL FINAL

Anabella: ¿En qué proyectos estás trabajando ahora o en qué estás poniendo atención?

Germán: Bien. A ver, lo que estamos laburando ahora es parte de lo que yo vengo laburando en CONICET, que tiene que ver con expansión de frontera agraria. Hay estudiantes de grado que están trabajando su tesina también alrededor de esa temática. Es un trabajo plural inclusivo, plural en serio, porque estamos trabajando para identificar en el mapa distintas formas de expansión de la frontera agraria con productos derivados de imágenes satelitales y estamos tratando de construir trayectorias de uso del suelo, a nivel pixel, digamos, a nivel de cien metros cuadrados, con el objetivo de poder mapear y cuantificar cómo cambió en cada pixelito, en cada manzana, cómo cambió el uso del suelo en los últimos 10, 15 años.

A mí es un tema que me resulta apasionante, porque es posible establecer algunas analogías “metodológicas” entre el tipo de dato que usamos (trayectorias de

uso del suelo) con cuestiones vinculadas a trayectorias vitales. En efecto, si abstraemos el dominio específico (suelo, trayectorias de vida, secuencias de ADN) lo cierto es que estamos lidiando con datos de secuencias. Entonces, allí se abre la posibilidad de usar una serie de herramientas de análisis de secuencias sumamente útiles para construir, por ejemplo, tipologías de cambios en el uso del suelo²². Es más, esas herramientas son análogas en ese mismo sentido a buena parte de las técnicas de PLN. Hay una homología “metodológica” allí que me resulta muy interesante. Entonces, estamos explorando un poquito algunas de esas herramientas y algunas fuentes en esa dirección.

Y después estamos empezando un poco a laburar ahora sí con LLMs, estamos empezando a pensar cómo empezamos con un *Silicon Sampling*²³ -un “muestreo de silicio”, esto es, de alguna manera, aprovecharse un poco de los sesgos que efectivamente tienen los LLMs para plantear preguntas interesantes de investigación. Hay una serie de trabajos que están explorando esta cuestión. Esto, en el fondo no deja de hacer una simulación. Si yo le pido a un LLM que asuma un cierto rol y le hago algunas preguntas, en las cuales yo conozco el resultado en la población general puedo comparar ambos resultados. Estamos pensando un poco, en cómo explorarlos un poco más para el contexto latinoamericano y argentino en principio. Hay muchos trabajos sobre Estados Unidos, como se imaginarán, pero para mí hay una serie de cuestiones interesantes que tienen que ver con explorar perfiles poco alcanzados por las encuestas. Te diría que en principio, eso son los laburos que estamos trabajando ahora sí de investigación.

Seguimos trabajando con la diplomatura, con una especie de reingeniería de algunos contenidos y pensando en cómo actualizar algunos contenidos que nos quedaron medio obsoletos. Sobre todo el módulo de NLP. Este año vamos a hacer una clase de LLM

22 Un primer resultado puede leerse en Rosati, G. (2023). “Analizando trayectorias de uso del suelo. Una propuesta de clusterización”, *Geograficando*, vol. 19, n°1, disponible en <https://www.geograficando.fahce.unlp.edu.ar/article/view/geo0130>.

23 Te invitamos a explorar este artículo: <https://www.cambridge.org/core/journals/political-analysis/article/out-of-one-many-using-language-models-to-simulate-human-samples/035D7C8A55B237942FB6DBAD7CAA4E49>

Francisco: Para cerrar, ahora sí ¿Qué futuro le ves a la sociología?

Germán: Como mencionábamos antes, hay una cuestión que tiene que ver con la matriculación en las carreras de grado. Me da la sensación de que es algo un poco general. No creo que sea un problema localizado en la UBA, pasa lo mismo en la UNSAM, aunque aparentemente no fue tan grave. Pero es bastante claro que hay un ataque muy fuerte sobre las Ciencias Socia-

les. Es evidente, entonces, que no se trata de una etapa expansiva.

Pero quizás este contexto sea una oportunidad para repensar ciertas cuestiones. Puede ser un contexto más o menos propicio como para volver sobre algunas prácticas que tenemos naturalizadas y que, en última instancia tienen que ver con todo lo que estuvimos hablando: cierta forma de investigar, cierta pedagogía, cierto tipo de perfil que se está generando en las Ciencias Sociales.